The Investigation of Alternative Weighting Approaches to adjust for Non-response in Longitudinal Surveys

Husam Eldin Sadig Ahmed Sadig

A thesis submitted for the degree of Doctor of Philosophy in

Survey Methodology

Institute for Social and Economic Research

University of Essex

June 2015

Declaration

This thesis has not been submitted, neither partially nor wholly, to any university for another degree. I prepared the data, conducted the analysis and wrote the thesis.

An earlier version of chapter 1 was presented in the Second Italian Conference on Survey Methodology (ITACOSM), 2011 under the name “Non-response weight adjustment in longitudinal surveys”. The paper was published in *Survey Research Methods and Applications*, p. 207 (conference proceedings) (see appendix A.1).

Acknowledgements

I greatly appreciate the support and inspiration of my Parents who had faith in me, and whose encouragement had confidence-boosting and lasting effect.

I would like to express the deepest appreciation to my principal supervisor, Professor Peter Lynn, who has the approach and attitude of a genius. He continually and thoroughly provided support both at the academic and social levels. Without his guidance, respect to the abilities of others and his unique supervision style, learning would have been difficult. I thank my second supervisor Dr. Olena Kaminska, who conveys a spirit of energy with regard to supervision and enthusiasm with regard to research. Her support is really acknowledged.

I also thank the Institute for Social and Economic Research (ISER) for providing such a stimulating environment for research. I appreciate all the support I received from my colleagues at ISER (students and researchers), particularly Chetti Nicoleti, Alita Nandi, Francisco Perales-Perez, Paul Groves, Janice Webb, Yadira Diaz, Maria Iacovou, And Cara Booker.

In addition, a big thank-you to the UK Economic and Social Research Council (ESRC) for funding two years of my PhD. Without the ESRC financial support, this project would have not been possible.

Finally, I am grateful to my friends in the city of Manchester who I turned to for advice and support during the hard times along the way of my study programme, especially Ahmed Humaidan, Walid Mohammed, Tareq Rahamtallah, Abdullah Musa and Mubarak Sherif.

Summary

To reduce bias in survey estimates, most longitudinal survey organisations, nowadays, prepare and include sets of weights in public use data files for use by analysts. Aside from correcting for non-coverage, the weights are usually designed to reflect the sample design as well as to correct for non-response error by combining design weights and non-response weight adjustments together.

With regard to non-response weights, many longitudinal surveys implement similar strategies (referred to as the standard weighting approach in this thesis) to create them. This approach is based upon a weighting model where: response is defined as responding at all conducted waves; all sample members whose eligibility is unknown are assumed as eligible and the model is estimated by using generic weighting variables and all sample members for which data are available on the weighting variables. However, there are several issues in longitudinal surveys that raise concerns regarding using this approach of weighting.

In particular, this thesis is concerned with three challenging issues: *non-monotonic response pattern* which results in a large number of combinations of waves at which sample members could respond, and hence weights that result from an approach such as the one in question, which defines response as responding at all the conducted waves may not be appropriate for the analysis of data from a wave-combination that does not include all waves; *unknown eligibility over time* leads to including a proportion of ineligible units in the weights' calculation (if they are assumed to be eligible as in the standard approach) which may result in biased estimates unless the actual ineligible units amongst units of

unknown eligibility are excluded; and *the choice of the best covariates for the weighting model* which may differ considerably across different subgroups of respondents in the same sample. In the standard approach only generic weighting variables are used in the weighting model, as all sample members are used in the estimation. Meanwhile, some variables, which may not be significant in predicting response for the whole sample, could be important in predicting the response in some subgroups.

In this thesis, I provide three alternative approaches (each deals with one of the raised issues) for non-response weighting.

I investigate each of the proposed approaches by incorporating relevant weight adjustments, as well as weights from the standard weighting approach, in a longitudinal multivariate analysis. I test the impact of weights from each alternative approach on estimates by comparing the resultant estimates with estimates resulting from the standard approach.

I use data from the British Household Panel Survey (BHPS) to carry out the investigation.

The findings suggest that the standard and alternative approaches, all help similarly in reducing non-response error. However, the standard approach may fail in tackling the effect of non-response in some estimates, as it does not take into account the three raised issues in the weighting of longitudinal data. In contrast, since they deal with the three issues under investigation (separately), the alternative approaches seem to handle non-response even in estimates that are not affected by the standard weighting approach.

Contents

List of tables and figures

(1) Tables

Chapter 1 tables

Chapter 2 tables

Chapter 3 tables

Appendix tables

(2) Figures

Chapter 1 figures

Chapter 2 figures

Introduction

About longitudinal surveys

Survey research represents one of the most important areas of measurement in applied social and economic research. This is particularly so in the last 60-70 years when household surveys (cross-sectional and longitudinal) have become a key source of data on social phenomena. A longitudinal survey, however, may be a more complex survey design than cross-sectional surveys, but it certainly offers several analytical advantages.

Cross-sectional surveys are conducted at a single time point. Thus, they are relatively less expensive and take less time to conduct compared to cross-sectional surveys. The data that are collected in a cross-sectional survey may provide an opportunity to analyse many substantive outcomes, and can be helpful to achieve several objectives (e.g. public health planning). However, as they are conducted at a single point in time, cross-sectional surveys ignore the fact that the same sample units may provide different measurements on the same variables if a different time frame was chosen, and hence analysis of unit-level change is not possible in these surveys.

Longitudinal surveys, on the other hand, may be more expensive and difficult to conduct, but they can provide data on the same set of units for a number of time points (waves). This enables the production of population cross-sectional measures every time data are collected, but more importantly allows the analysis of unit-level change. Experts in surveys analyse the advantages and disadvantages of both longitudinal and cross-sectional surveys in different ways, but the most recent, and probably the most informative discussion, available in (Lynn, 2009). There are two main types of longitudinal surveys:

Cohort studies focus on a particular population, but they sample from a specific age cohort. Typically, the sample drawn for a cohort study is selected from a birth cohort of individuals who were born in a single week or a month in a given year. The cohort study then follows the lives of the individuals selected in the sample and interviews them at particular ages at regular or (often) irregular intervals to explore patterns in specific socio-economic phenomena such as health behaviour and family life. For example, a survey organisation may decide to follow the lives of a sample of new born children who will be born in a single week in the year 2020 to understand the factors associated with the change in their health at different ages.

Panel studies also follow the same sample units and attempt to collect data from them at every data collection point. However, a major distinction between panel studies and cohort studies lies in the way that they select their samples. While a birth cohort studies sample from a specific age cohort, panel studies typically target the entire age range in a given country to explore the dynamic of change (in a wide range of phenomena) experienced by the resident population in the country. For example, a survey organisation may randomly select individuals from randomly selected households in a resident population in a given country, and interview them about various social phenomena. Every year, the same individuals can be contacted and asked similar questions and the reasons for any change. Panel studies tend to have more frequent data collection points (waves) compared to cohort studies. However, this can yield extremely specific and useful explanations of social phenomena. Thus, as they target larger populations and collect data more frequently than cohort studies, panel studies tend to be more complex and more difficult to conduct. As a result, they can suffer from more problems. The focus in this

research is on panel studies, in particular household panel studies, and specific type of problems (errors) that occur in these surveys which will be explained later.
In recent decades, the world has seen the execution of large household panel studies. Some of these studies implement the best procedures in the art of survey design that survey research has developed. Some of the major household panel studies in the world are:

The British Household Panel Survey (BHPS) conducted in Great Britain (1991-2008). The BHPS is a result of a proposal to the UK Economic and Social Research Council (ESRC) to establish an interdisciplinary research centre at the University of Essex (Lynn, 2006). More details on the BHPS design, sample and other features will be given later as this is the main data source for this research. In 2009, Understanding Society took over from the BHPS (the BHPS was incorporated into Understanding Society) as the new UK household longitudinal study. With a sample of 100,000 individuals, Understanding Society is currently (2015) the world's largest survey of its type.

The Panel Study of Income Dynamics (PSID) is the oldest longitudinal panel study. PSID started in the USA in 1968 and has been collecting measurements from the same sample ever since (Duncan *et al*, 2004).

The German Socio-economic Panel (GSOEP) is the household panel study of the population in Germany. It was started in 1984, and is conducted by the German Institute for Economic Research (DIW Berlin) (Kroh, 2009).

The Household, Income and Labour Dynamic in Australia (HILDA) started in 2001 as Australia's household-based longitudinal survey. HILDA pays more attention to family and household formation, income and work than other socio-economic topics (Summerfield, 2010).

The Swiss Household Panel (SHP) started in Switzerland in 1999. Based on the country's telephone directory, the survey covers individuals who are resident in private households in Switzerland who have a registered landline or mobile phone (Plaza, 2008).

Errors in surveys

The term *survey error* does not necessarily mean a 'mistake' in the linguistic sense it is rather a deviation from the desired outcome (Groves *et al*, 2004).

Surveys in general suffer from many types of error (explained in Groves *et al*, 2004 & Scheaffer et al, 1996). On the one hand, these errors may arise because the measurements collected from the set of sampled members that participated in the survey do not accurately reflect the underlying attributes that the survey was designed to study (errors of observations). In this case, the errors may result because: the survey fails to choose appropriate measures to represent the phenomena that are studied (validity); responses given to the survey questions may be incorrect (measurement error) either because the questionnaire used to collect the responses is not well designed or because the respondent intentionally provides incorrect responses; or in unfortunate situations the error may result during the process of editing the answers provided by survey respondents (processing error).

On the other hand, errors arise in surveys because only a sample from the population is intended for measurement (errors of non-observations). In this respect, errors may emerge because the sampling frame used to select the sample may itself suffer from some problems (coverage error). For instance, if the sampling frame does not include all units in the target population (non-coverage error), some units may not be subject to sampling, which may lead to misleading results if the non-covered units have different measures on the survey target variables than the units included in the sampling frame. Also, since not all units in the sampling frame are selected, the mechanism upon which sample members are included in the sample may, itself, result in some form of error (sampling error). For example, if the sample design does not allow some units in the sampling frame to be selected in the sample, and these units have different characteristics on the substantive survey variables than the rest of the units in the sampling frame, this may distort the results produced from the selected sample (sampling bias). Moreover, even for the sampled units, the obtained data may be incomplete (non-response error) or, in other words, a missing data problem might exist. The error that results from non-response could be one of the serious errors in survey research as will be explained in more detail in the next section. The subject of this thesis, however, is centred around dealing with this error at the analysis stage in household longitudinal surveys.

To tackle errors of non-observations (including non-response), a range of techniques may be adopted. Some of these techniques can be implemented before and at the data collection stage. Other techniques may be applied after (i.e. at the analysis stage). Methods that are used after the data collection are referred to as 'post-survey adjustments'. These methods (which will also be explained later) attempt to adjust the collected data to account for issues arising from non-observational errors. However, any post-survey adjustment method requires assumptions that may not necessarily hold, and in many cases may be difficult to verify. Thus, these methods are, themselves, subject to error. The error that may result from post-survey adjustments (adjustment error) is also classified as a non-observational error (Groves *et al*, 2004).

Non-response

The term 'non-response' is used to describe the situation when the survey organisation fails to collect responses from some of the sampled members who are eligible for the survey (Lynn, 2008; Bethlehem, 2009). A partial failure is referred to as *item non-response*, whereas a complete failure (i.e. no measurement is provided on any of the survey variables) is called *unit non-response* (de Leeuw *et al*, 2003 Lessler and Kalsbeek, 1992; Madow et al, 1983). This thesis is concerned with unit non-response, and the term 'non-response' may be used in this thesis to refer to unit non-response. Non-response occurs in every survey as it is very difficult to obtain all the required data from the selected sample. This includes even the most well designed surveys conducted by highly experienced survey organisations (Lynn, 2008).

Reasons for non-response

The propensity to respond to surveys differs across individuals. Some people are easy to approach, and may be also easy to obtain data from. Other people are not. Thus, there is a wide range of reasons for which non-response may arise. The literature on non-response generally distinguishes among five different reasons (see Lessler and Kalsbeek, 1992; Lepkowski and Couper, 2002; Lynn, 2008; Groves et al, 2004; Kish, 1965): (1) *failure to locate sample members*: it is sometimes impossible for the survey researcher to successfully locate sampled members if, for example, their address is incomplete, (2) *failure to deliver the survey request to located sample members:* this refers to a situation where a sampled member is successfully located but the interviewer is unable reach them to conduct the interview because the sampled member is not available (e.g. not at home). Non-response that result from (1) and (2) is known as '*non-contact*', (3) *failure to gain cooperation from sample members:* this happens when a successfully contacted sampled member is unwilling to take part in the survey. This depends mainly on factors such as the nature of sample members; some individuals are more cooperative than others due to culture, social class and the demographic categories they belong to. This type of non-response is known as '*refusal*', (4) *incapacity or inability:* in this case, the approached sampled member may be willing to take part in the survey; however, he or she is unable due to illness, illiteracy or a language barrier, and (5) *loss of information:* this refers to the accidental loss of data after the data collection. For example, questionnaire forms might get lost in the post or destroyed if neglected.

Non-response that occurs through inability and loss of information generally only represents a small proportion of non-responding cases compared to non-contact and refusal. However, most of non-response in surveys nowadays is due to refusal rather than non-contact (Brick, 2013; Atrostic *et al*, 2001).

In order to better understand the factors associated with the different reasons for non-response, survey researchers usually analyse non-response, based on non-response correlates, into its main sources: refusal and non-contact.

Refusal

In sample surveys, correlates of refusal can be divided into five groups (Groves and Couper, 1998). These are: *Social Environmental Factors* which include factors such as urbanity and crime rates in the neighbourhood; *Respondent Factors* include demographic characteristics, household composition and personal attitude; *Survey Factors* such as the survey design, survey sponsor, survey topic and data collection mode; *Interviewer Factors* are experience, gender, race, age and attitude towards the survey task; and *Interaction between the Interviewer and the Respondents* is the conversation between them which takes place during the interview.

A few attributes have been found to be highly correlated with refusals (Groves and Couper, 1998). These are:

Gender: females usually have lower refusal rates compared to males, since women are more inclined to engage in conversations than men (Smith, 1984).

Urbanity: studies have shown that those who live in urban areas are more likely to refuse than those in non-urban areas (Steeth, 1981). This was explained by Groves and Couper (1998) as avoiding contact with others as a fear of crime since crime rates are higher in urban areas than in rural areas.

Single-person household: some individuals who live alone may be socially isolated. Socially isolated individuals may not feel obligated by civic duty to cooperate with surveys, and may not be willing to be found by the interviewer (Brehm, 1993).

Survey topic: individuals who are interested in the survey topic are more likely to cooperate than others (Groves, Presser and Dipko, 2004). For instance voters and those who are interested in politics are more likely to cooperate in election surveys (Couper, 1997; Brehm, 1993). On the other hand, sensitive topics (e.g. self opinion about same sex marriage) are more associated with refusal (Lynn, 2008).

Survey sponsor: sponsors that have an authoritative nature generally have high cooperation rates. For example, government surveys generate fewer refusals than academic surveys, which in turn generate fewer refusals than commercial surveys (Groves and Couper, 1998).

Survey design: surveys that implement strategies in their design to encourage sample members to cooperate tend to result in high cooperation rates. For instance, individuals may be less inclined to refuse if the survey offers a reward, like a gift, for those who participate (Laurie and Lynn, 2009).

Non-contact

In household surveys, where interviewers attempt to contact sample members at their homes, non-contact is primarily caused by the fact that sample members are not at home at the time of the contact attempt. This is usually referred to as *At-home pattern*. In addition, contact may not be established with sample members if their homes have restrictive access procedures that impede interviewers from making the contact (e.g. a gated house). This is called *Access impediments*. Both *At-home pattern* and *Access impediments* are highly associated with two issues: type of sample unit and mode of data collection.

With regard to *At-home pattern*, for example, males, employed adults, and young single individuals spend more time away from home and are therefore difficult to contact (Groves and Couper, 1998). Also, in surveys that interview respondents over the phone, non-contact due to at-home pattern depends on the time of the call. Call attempts that are made during evenings and weekends could be more successful than calls made during weekdays (Bates, 2003; Brick *et al*, 1996).

As for *Access impediments*, this prevents contact in face-to-face interviews if respondents' homes are based in, for example, locked blocks of flats, gated houses, buildings with security devices that limit contact with residents (Groves and Couper, 1998). Such Access impediments are common in high-security neighbourhoods. In telephone surveys, Access impediments may be in a form of a device such as answering machines or caller identity device (Tuckel and O'Neill, 1995).

Analysing non-response based on non-contact and refusal is useful. It improves our understanding of non-response based on two different causes of the phenomenon. Nonetheless, in my opinion, restricting the analysis of non-response to non-contact and refusal tells us little about the actual reasons for non-response. In fact it is not refusal or non-contact per se that causes non-response. The actual reasons for non-response are likely to be related to the circumstances of sample members at the time that the survey request is made. Such factors, which result in either refusal or non-contact, are not usually measured by the survey. In turn, with non-contact and refusal we can only get a one-level explanation which does not necessarily reflect the concrete reasons for non-response. In order to develop a deeper understanding of non-response, we may need to break the analysis down into another level and inspect why those who, for example, were not contacted were difficult to contact.

For instance, some people may not be contacted because they are searching for a job outside their homes, other people may not be contacted because they have left the country where the survey is conducted (by which they may not be part of the target population). These two explanations may lead to completely different consequences of non-response. Therefore, they can be useful in determining circumstances when we should worry about non-response and when we should not. Moreover, one can immediately notice that explanations of non-response on two different levels (e.g. level 1: non-contact, and level 2: not available because looking for a job) do not contradict, they rather complement.

Effect of non-response

Non-response leads to one of two problems: (a) if many sample members do not participate in the survey, the sample size that one had hoped for at the design stage will be reduced. Thus, estimates derived from the smaller sample will be less precise. This is however a minor problem, as the sample size can be set to a required achieved sample according to a predicted non-response level (Lynn, 1996); (b) if many sample members do not respond to the survey, and those who do not respond have different values from respondents on variables that are components of the survey statistics, estimates based solely on information from respondents can be biased. The combination of (a) and (b), which increases the Mean Square Error (MSE) of survey estimates, is referred to as 'non-response error'[1]. However, the bias is the dominant component, and is the reason that concerns are raised about non-response error (Lynn, 1996).

Non-response bias is a deviation in a statistic that is estimated on the set of responding sample from one that estimated on a full sample. This deviation results as a consequence of a systematic distortion of the response process. For estimates such as the mean of a

[1]In the same survey, non-response error varies across estimates. For example, if respondents are systematically different from non-respondents on a variable 'y' but similar to non-respondents on another variable 'x', estimates derived from 'x' will be less affected by non-response error.

target variable Y (represented by the corresponding sample statistics $\bar{y}$) non-response error is the deviation in the value of $\bar{y}$ for respondents from the value of $\bar{y}$ for the full sample. Taking the non-respondents into account, this can be expressed by equation (1).

$$\bar{y}_r - \bar{y}_n = \frac{m}{n} * (\bar{y}_r - \bar{y}_m) \tag{1}$$

Where *n* is the selected sample size, *m* is the number of non-respondents and *r* denotes the respondents. Thus, $\bar{y}_r$ is the value of $\bar{y}$ for respondents; $\bar{y}_n$ is the value of $\bar{y}$ for the full sample; and $\bar{y}_m$ is the value of $\bar{y}$ for non-respondents. The left hand side of the equation represents non-response error, which is expressed in the right hand side as a product of the non-response rate and the difference between respondents and non-respondents in the estimate in question.

Link between response rate and non-response bias

In many surveys, one of the main concerns is to increase the response rate. Increasing the response rate is desirable since it automatically decreases non-response rate and hence may minimize the likelihood of the bias linked to non-response. This indicates that non-response bias may strike surveys with low response rate more than surveys with high response rate. However, the magnitude of the response rate does not provide information about the existence or the size of non-response bias (Groves, 2006; Groves and Peytcheva, 2008).

It is possible to have little bias even if the response rate is low if respondents and non-respondents do not differ largely in terms of what is being estimated (i.e. if $(\bar{y}_r - \bar{y}_m)$ is small). In fact, with a low response rate non-response bias might not even exist if respondents and non-respondents are very similar in all the characteristics related to all the survey key variables. In return, it is also possible to have high bias with high response rate if respondents and non-respondents differ greatly in the characteristics in question (i.e. if $(\bar{y}_r - \bar{y}_m)$is large). Thus, the key component of non-response error is $(\bar{y}_r - \bar{y}_m)$ (i.e. the difference between respondents and non-respondents in the estimated statistics), and not $(\frac{m}{n})$ (i.e. non-response rate).

Survey researchers sometimes attempt to examine the existence of non-response bias, or in other words, estimate $(\bar{y}_r - \bar{y}_m)$. Nonetheless, it is impossible to discover differences between respondents and non-respondents using data with respect to a survey target variable Y, because its measurements are only available for respondents. In other words, $y_m$is not observed. However, auxiliary data that are available for both respondents and non-respondents can be used to inspect the differences between respondents and non-respondents[2], but may not necessarily provide good estimates of the bias with respect to the substantive survey variables. This is because it is sometimes difficult to assert that respondents and non-respondents are different with respect to a target variable even if they differ in terms of other auxiliary characteristics.

There seems to exist a gap with regard to a method that can detect the existence of non-response bias accurately. In return, the non-response rate $(\frac{m}{n})$ remains an important indicator, in this respect, which survey researchers try to minimise as protection against bias. Therefore, in recent years, the link between response rate and non-response bias has been an interesting topic among survey researchers. Although many survey researchers

[2] Groves (2006) provides methods for assessing non-response bias and reviews their strengths and their weaknesses.

stress the importance of endeavouring to increase the response rate (for example Alreck and Settle, 1995), several studies on the other hand found that changes in response rates may not necessarily have an impact on the survey estimates (Curtin, Presser and Singer, 2000; Merkle and Edelman, 2002).

Non-response in longitudinal surveys

In recent decades, more attention has been paid to longitudinal surveys and the way that they are carried out. This has resulted in a rapid improvement in the design and execution of longitudinal surveys. For example, the involvement of computer technology in data collection has resulted in a reduction in the cost of all surveys and also improved the quality of the data (Bethlehem, 2009). However, new problems also emerged. One of these is the increase in non-response rates. Despite the effort that survey organisations implement nowadays to improve the survey design in order to achieve the highest possible response rate, non-response rates are still rising in most longitudinal surveys in the world (Watson and Wooden, 2009). As a result, the implementation of post-survey adjustments is becoming more popular amongst survey researches.

A distinctive feature of household longitudinal surveys is that they collect observations from individuals on multiple occasions. This design involves following individuals over time and continuing to collect data from them. However, respondents might not be available to participate in the survey every time data are collected. Therefore, non-response can occur for a number of reasons that result in non-contact or refusal (Lepkowski & Couper, 2002). For example, if some sample members change their address without informing the survey organisation, this might result in non-response; also, some respondents may refuse to respond at some point although they have participated in previous waves. Thus, the complexity of longitudinal surveys turns non-response into a dynamic event that accumulates over time when further waves are conducted (Watson and Wooden, 2009). This may be a dilemma for the survey organisation, especially if respondents were chosen via a probability sample design since they cannot be replaced. Logical questions that the survey organisation may face in these circumstances are: what advantage a probability sample has if it will (over time) suffer from high non-response? Would a non-probability sample be a better option even though it may not be representative of the population of interest?

Non-response in longitudinal surveys can be in one of two forms: (1) *Wave non-response* refers to the process where a sample member is absent from the survey for at least one wave but returns to the survey in a later wave. (2) *Attrition* on the other hand occurs when a respondent participates in the survey for one or more waves but permanently stops participating at some point during the survey course. Although the former is not trivial, survey researchers are more concerned about the latter for at least three reasons: (a) more information is lost in the case of attrition; (b) the potential bias caused by non-response is more likely to occur (Chang, 2010); (c) any observed information collected in earlier waves become weak predictors as more waves are added (Chang, 2010). In surveys with an indefinite number of waves, it is always difficult to distinguish wave non-response from attrition as it is controlled by the respondents' behaviour in the future. In contrast, the point of attrition can be identified in finite length surveys when no further waves are conducted (Uhrig, 2008).

Causes of attrition and wave non-response

Attrition and wave non-response are special cases of non-response occurring in panel studies. The causes of non-response in longitudinal surveys may be similar to those in cross-sectional surveys but somewhat different in at least two ways: (a) in a longitudinal survey respondents are burdened by a constant long-term commitment to responding; and (b) changes occurring during gaps of data collection points may have a big effect on the response process.

Therefore, some reasons for non-response may be specific to longitudinal studies. For example:

(1) *Failure to update contact information:* If survey participants move houses between waves without informing the survey organisation, it could be too expensive if not impossible to track them and failure to do so will directly result in non-contact.

(2) *Loss of interest:* Although some respondents may be interested in taking part in the survey at the start, their level of willingness to continue giving data at every wave is a function of the survey organisation's effort to maintain their interest level (e.g. use of incentive). Failure to retain participation interest results in refusal.

(3) *Changes in health condition*: Longitudinal surveys are conducted over a long period of time. Over this period, some respondents might suffer from a bad health condition leading them to have to drop out.

(4) *Technical issues related to data collection strategy*: Survey organisations may be forced sometimes to adopt changes in the data collection strategy. For example, a failure

in maintaining the interviewer from the last wave may affect the propensity to respond in the current wave. This is particularly common when the new interviewer is less experienced. Although there is no evidence that the change of interviewer results in non-response, recent research shows that interviewer continuity is associated with low propensity for refusal (Lynn, Kaminska and Goldstein, 2014).

The effect of non-response in the longitudinal context

The potential bias and smaller sample size that result as a consequence of non-response may be more problematic in longitudinal surveys for at least two reasons:

(1) The original sample may suffer from a monotonic decrease in its size. As a result, after a large number of waves, the survey organisation might end up with a relatively small sample that is incapable of producing precise estimates. In cross-sectional surveys, I mentioned that this is not a major problem since a required achieved sample size can be set according to a predicted non-response level. In contrast, in longitudinal surveys, even if the survey organisation invested in a very large sample size the reduction of the sample size may still be a problem in the long term, particularly in surveys of indefinite length. This distinction shows that the decrease in the sample size due to non-response is more problematic in longitudinal surveys, and in fact it is a typical feature of panel data. Therefore, it would be very wise for survey organisations to establish future plans at the design stage to deal with this problem in order to increase the size of the achieved sample and hence ameliorate the precision of the survey estimates.

(2) The process in which respondents drop out of the study may not be random (Fitzgerald, Gottschalk and Moffitt, 1998; Watson, 2003) and it is plausible to assume

that some of the drop outs are because of the topics covered by the survey. For example, respondents may participate in one or two waves and then decide to drop out after they have discovered what the survey is about. If these drop outs are different from respondents in terms of what the survey is measuring, the sample becomes progressively unrepresentative of the population as more waves are conducted. Consequently, estimates derived from the achieved sample will be biased.

Methods for dealing with non-response

There are two broad categories of methods that are used to deal with the non-response problem. Before describing these categories, it is important to point out that it is a good idea to combine methods from both categories in order to tackle non-response effectively. The first group of methods is concerned with minimising non-response when collecting the data (Lynn, 1996; Stoop et al, 2010). In this regard, survey organisations may incorporate a mixture of techniques in the survey design in trying to decrease non-response rate to its minimum. Groves *et al* (2004) provide a wide range of these. Examples of design features that may reduce non-response are:

Increased number of contact attempts: it is well documented in the literature of non-response that the larger the number of contact attempts, the higher the likelihood of successful contact (Goyder, 1985).

Long data collection period: long data collection periods provide a higher chance for successfully delivering the survey request to a larger group of sample members.

Reduced number of sample members assigned to interviewers: if a large number of cases are assigned to an interviewer, less effort may be given to convince some cases to respond.

Pre-notification letters: unexpected calls, or visits from ‘strangers’ (the interviewers) may cause some sample members to refuse.

Use of incentives: offering a gift or money to sample members in return of their participation may encourage many people to take part in the survey (Laurie and Lynn, 2009).

Mixed-modes of data collection: some sample members can only be contacted through specific mode (e.g. face to face), other sample members may be contactable via a number of different modes (e.g. telephone, mail and web). Using a mixed-mode design, thus, increases the likelihood of contacting a larger number of sample members.

Interviewer/ household matching: some interviewers may have characteristics that more ‘acceptable’ to certain groups of sample members (e.g. female interviewers may be preferable to old women who are living alone). If these can be identified, it is wise to assign sample members to interviewers in a way that improves the likelihood of trust, and hence increase the response tendency.

Although these methods are very useful in increasing the response rate, still, it is impossible to obtain 100% response rate, especially in surveys targeting households and individuals.

Thus, a second and important group of methods deals with non-response at the analysis stage. As mentioned before, these methods are referred to as 'post-survey adjustments'. The rationale behind these methods is based on the idea of adjusting the distribution of the responding sample (achieved) in a way that makes it similar to the distribution of the target population. As a result, those who are missing from the sample will be compensated for through this adjustment. Consequently, estimates produced from the adjusted sample may be less biased.

Before introducing the types of post-survey adjustments that are used to deal with non-response, it is important to distinguish between different types of 'missing data mechanisms'. A missing data mechanism is the process that generates the missingness. The choice of specific post-survey adjustment method to deal with missing data relies on our expectations on the missing data mechanism.

Missing data mechanisms

There are three main types of missing data mechanisms in the literature that can be distinguished (given by Rubin, 1976; Little and Rubin, 1987; 2002; Allison, 2000; Bethlehem, 2009). If Y represents a substantive survey variable for which some values are missing for some of the sample members; X represents a set of auxiliary variables that is fully observed for all sample members; Z is a variable that is external to the survey and is uncorrelated with X and Y; and R indicates whether values of Y are observed or not, then:

Missing completely at random (MCAR) is a situation where the missingness is caused solely by the outside phenomenon Z (i.e. R and Z are correlated). In this case, estimates derived from Y will not be biased and post-survey adjustments are not necessary.

Missing at random (MAR) refers to the situation where the missingness depends on *Z,* but it may also depend on *X* so that there is an indirect relationship between *Y* and *R* (i.e. the relationship between *Y* and *R* is conditional on *X*). If this is the case, the missingness may cause bias to estimates derived from *Y*, but fortunately, by using *X*, a number of post-survey adjustments may be used to adjust for this bias. If MAR holds and the parameters governing the missing data mechanism are distinct from the parameters to be estimated, the missing data mechanism is said to be 'ignorable'. It is common to use the terms *MAR* and *ignorable missingness* interchangeably because, in practise, the parameters to be estimated are likely to be distinct from the parameters governing the missing data mechanism (Allison, 2000).

Not missing at random (NMAR) is the case where there is a direct relationship between *Y* and *R* (and may be also between *X* and *R*; and *Z* and *R*). In other words, the missingness is caused by the survey variables. In this situation, estimates derived from *Y* will be biased. Unfortunately, this will limit the choice of the post-survey adjustments that can be used to deal with the problem, as some methods cannot help reducing the bias in this case.

In practise, however, survey researchers will not know which of the three missing data mechanisms applies to the data. Because, such knowledge requires full measurements on the selected sample in terms of the survey key variables, which in turn makes addressing the problem of missing data unnecessary in the first place. Thus, survey researchers have to assume one of the three missing data patterns. MAR is the most assumed missing data mechanism, as it allows the implementation of a wider range of post-survey adjustments.

Although the concept of a single missing data mechanism is usually presented in the literature on missing data as operating at the level of the sample, in my opinion, the missingness mechanism in a substantive variable *Y* may very well be a combination of two or even the three missing data mechanisms (MCAR, MAR and NMAR). This may occur if there is no homogeneity between the non-responding sample members with respect to the actual reasons of non-response. In this case, for some sample members, data may be missing on *Y* under MAR, and for others may be under NMAR. Thus, assuming a single missingness mechanism may not necessarily reflect the actual missing pattern in the data. Nonetheless, it is helpful in terms of deciding between classes of missing data adjustments. One of the areas that is not fully understood is whether a multiple missingness mechanism rather than a single missingness mechanism affects the adjustment of missing data.

Post-survey adjustments

Post-survey adjustments are a class of methods that are used to tackle errors of non-observations. The focus here is on non-response error. There are a number of post-survey adjustments that can be used to deal with the non-response problem. In this section, I discuss the most common methods. These are: Non-response Weighting; Post-stratification; Calibration; Raking; Multiple Imputation; and the Selection Model Approach. Some of these methods, such as post-stratification, are not primarily used to adjust for non-response (as will be explained next); still, they may relatively reduce non-response bias if they meet certain conditions as will be explained.

It is important to point out that all these methods rely on, and use a set of auxiliary variables '*X*' to deal with non-response in the set of survey target variables '*Y*'. Some methods require *X* to be known for all cases in the target population, or otherwise the sampling frame (calibration; post-stratification; and raking). Other methods only require measurements on *X* for the selected sample (non-response weighting; multiple imputation; and the selection model approach). The important issue here is that *X* needs to be observed for both respondents and non-respondents to be able to deal with non-response in one way or another. Also, to successfully adjust the responding sample to reduce non-response error, *X* needs to be associated with the response propensity. In the case of longitudinal surveys, substantive survey variables that were observed for all sample members from previous waves may also be included in *X*. Chapter 1 in this thesis provides details about the types and sources of auxiliary variables that can be used in this regard.

In this section we provide a brief overview of the above post-survey adjustments.

Weighting for non-response is a technique that assigns numerical values (weights) to the responding sample units, in order to modify them to also represent non-responding sample units (Lynn, 2005). As a result, it is hoped that the weighted distribution of the responding sample will be similar to that of the selected sample. More details on non-response weighting, including the construction of the weights, will be discussed in the next sections as this is the subject of this thesis. However, the term 'weighting' will be used here as shorthand for non-response weighting.

Calibration is a method that assigns values (also called weights) to respondents so that known parameters of the auxiliary variables *X*, either from the population or another survey, can be reproduced (Sikkel, Hox and de Leeuw, 2009; Särndal and Lundström, 2005). This procedure usually results in estimates with smaller standard errors. If the auxiliary variables used in calibration distinguish response from non-response (i.e. are correlated with the response propensity), non-response error can also be reduced.

Post-stratification also assigns values to respondents so that their sums are equal to known population totals for certain sub-groups of the population (Biemer and Christ, 2008). For example, if the population totals of subgroups defined by gender are known (maybe from the sampling frame or other external source), post-stratification assigns weights to respondents so that their distribution by the defined subgroups is the same as the known population distribution. In this respect, post-stratification can also be classified as a calibration method. The difference is that in calibration the known subgroups totals may not necessarily be from the population, they could be from another survey. Post-stratification is used primarily to correct for non-coverage error, and to reduce the variance of survey estimates. It is not typically used to deal with non-response bias. However, if the auxiliary variables that form the subgroups in the post-stratification are powerful predictors of the response probability, post-stratification may also reduce non-response bias.

Raking is an extension to post-stratification. It is a process that performs multidimensional post-stratification (Cervantes and Brick, 2008). It assigns values to respondents in order to match known population distributions in a number of auxiliary variables (dimensions). Raking repeats this process a number of times until accepted (tolerable) distributions are met. It is, thus, different from post-stratification by the fact that it does not reproduce the exact population distributions on the auxiliary variables. Another difference between raking and post-stratification is that, unlike post-stratification, in raking, the joint distribution of the auxiliary variables (cross-tabulation) does not need to be known. Instead, raking can be used, if only the marginal distributions of two or more auxiliary variables are known.

Multiple Imputation (MI) is different from single imputation (SI). The latter produces one synthetic value to replace a missing value in a target variable *Y*. This can be deterministic if, for example, the missing values in a variable replaced by the mean value of the variable; or random if the imputed values are selected randomly from the available values of the variable being imputed[3]. Bethlehem (2009) provides a rich discussion for a range of different SI methods. These are not discussed here as our focus is on MI. However, there are two major disadvantages of SI that can be mentioned (indicated by de Leeuw *et al*, 2003): 1) using the observed data to impute the missing values emphasizes the structure of the observed data in the imputed data set; and 2) analysing the imputed data set involves using a spuriously large number of cases which may lead to biased significance tests.

MI, on the other hand, may solve the problems of SI. MI produces a set of synthetic values to replace a missing value. The method originated in the 1970s in application to non-response (Rubin, 1976). The ordinary concept of MI (proposed by Rubin, 1987) is based on three steps: impute the missing values in the data *m* times (results in *m* complete

[3] Deterministic imputation results in the same value if the imputation is repeated; whereas in random imputation the imputed value may change if the imputation is repeated.

data sets); perform the desired analysis (*m* times) on each of the complete data (imputed) sets; and combine the results obtained from the *m-time* repeated analysis into a single result. Although analysing the data *m* times may seem inconvenient, it is not difficult, especially with the existence of a number of powerful conventional software nowadays. What might be difficult is the generation of *m* data sets in an appropriate manner (de Leeuw *et al*, 2003). In MI, the imputed values must include an error term from an appropriate distribution (generally, the models used for the data generation should include variables that predict either the missingness or the outcome variable). This solves the problem of emphasizing the existing structure in the data. Also, analysing *m* data sets and combining the resultant estimates into an overall estimate resolves the problem of the biased significance tests.

Over time, MI went through remarkable improvements to develop imputation models which use variables that predict both the missingness and the outcome of interest (Schafer, 1997) which results in a more efficient analysis. In recent years, a number of studies have demonstrated that MI can also be incorporated in dealing with non-response in substantive longitudinal data analyses (Goldstein, 2009; Carpenter and Plewis, 2011; Plewis, 2011). One important difference between imputation and calibration-based methods is that imputation attempts to produce a distribution that resembles the true distribution of the imputed variable, which is not required by calibration-based methods.

All of the above post-survey adjustments assume MAR (although MI can also work under NMAR if the imputation model can correctly specify the missingness). In circumstances where survey researchers have reasons to believe that MAR does not hold, the missing data mechanism is not ignorable, and valid estimation may require modelling the missingness as part of the estimation process. MI can produce valid estimates, in this case, if the model for missingness is correctly specified (Allison, 2000). However, these situations are fraught with difficulty. Because the very data that suffer from missingness cannot support the specification of an appropriate model that correctly predict the missingness.

Another well-known method that is used for handling missing data when MAR is violated is the selection model approach.

Selection Model Approach (SMA) is a method that assumes NMAR. As we mentioned, if data are not missing at random, there are no simple solutions. A specific model for missingness must be hypothesized. The SMA then postulates a model that links the missingness to the distribution of the outcome variable (Heckman, 1979; Hausman and Wise, 1979). In other words, it jointly models the substantive model of interest and the probability that the outcome variable is observed.

However, there are a number of drawbacks in the SMA. Aside from the fact that there is no information in the data to help chose an appropriate model, there is no statistics that can show how well a chosen model fits the data and the results are often sensitive to the choice of the model (Little and Rubin, 2002). Furthermore, applying this method requires the availability of variables that are not correlated with substantive outcome. Fully observed instrumental variables (from previous waves in case of longitudinal data) that vary between units and predict the missingness may be good candidates. However, such variables, sometimes, are difficult to find as most available survey variables are respondents' personal characteristics, in which case they are likely to be related to the substantive outcome. Also, other variables such as characteristics of interviewers and interview condition have little variation across units (Fitzgerald, Gottschalk and Moffitt, 1998). Recent studies that applied SMA to data from longitudinal surveys are Carpenter and Plewis (2011) and Plewis (2011).

As mentioned earlier, adjusting for missing data appropriately depends on the missingness mechanism and the method that is used as a post-survey adjustment. However, a simulation study by Collins, Schafer and Kam (2001) reports that the MAR-based methods result in little bias in estimates even when the missingness is NMAR. The only exception is that when some of the causes of missingness that are not included in the adjustment are strongly correlated with the substantive variable Y (with a correlation coefficient greater than 0.4).

Weighting

Weighting is an adjustment which is implemented at the stage of analysing the data. It is applied to compensate for the units missing from the selected sample (non-responders). It adjusts the responding sample so that its distribution is the same as the selected sample, and hence produces unbiased estimates.

In weighting we calculate numerical entities (weights) that represent the influence of survey respondents on estimates. When constructing a survey-based estimate, weighting assigns the calculated weights to respondents as their contribution to the estimate in question (Lynn, 2005). The weight of a respondent can be interpreted as the number of individuals in the target population that are represented by the respondent.

Aside from non-response weighting and calibration-based weights, there is another type of weighting that is usually used in conjunction with non-response weighting. This weighting is used to reflect the selection procedure, when the sample is selected with unequal probabilities of selection. Weight values that result from this type of weighting are referred to as 'Design weights'. In practise, the design weights are created first before adjusted with non-response weights. In this respect, the final analysis weight used to adjust for non-response is a combination of the design weight and non-response weight. Thus, it may be important to discuss the role of the design weights and their construction before explaining how non-response weights are created.

The design weights

The design weights are used to correct for the *unequal probabilities of selection*. This occurs when some of the units in the sampling frame have a different chance of being in the selected sample than other units. If a sample has been selected with unequal probabilities, estimates such as the unweighted sample mean are biased (Horvitz and Tompson, 1952). For example, consider a sample design that aims at randomly selecting one adult from each of H households. In this example the chance of an adult being selected from household h depends on the number of adults in this household. In other words, the probability of selection increases as the number of adults in the household decreases. Thus, ignoring the fact that the selection probabilities are different will result in bias in estimates due to an over-representation of households with fewer adults. This can be avoided if a correction is implemented to balance the probabilities of selection. This correction is the design weight: it adds more value to the cases whose probability of selection is low to represent more cases of their category, and decreases the value of the cases whose probability of selection is high, in order to balance the sample. Therefore, the design weight for a given unit in the sample is the inverse of the selection probability for this unit. Thus, calculating the design weights is fairly simple. It only requires knowledge of the selection probabilities for every unit in the sample. The calculation of the design weight is given by equation (2).

$$D_i = {p_i}^{-1} \tag{2}$$

Where D_i is the design weight for case i; and p_i is the probability of selection for unit i.

If sample units were selected using a simple random sampling method, p_i becomes constant. In this case, all sample units will have the same design weight which is the ratio of the number of units on the sampling frame to the number of units in the selected sample. Otherwise, the design weight must reflect the strategy of selection for each unit separately.

Constructing non-response weights

Although the rationale behind non-response weights is convincing and well established, there is no universally held protocol to compute them. Weights construction varies according to the differences in circumstances from sample to sample concerning the design and the availability of auxiliary information about the sample and the target population (GATS Sample Weights Manual, 2009). Thus, the actual stages for deriving the weights may vary from one survey to another. Therefore, the weights are usually created and released by the survey organisation. Nevertheless, there are general well-known steps to constructing the weights, to compensate for non-response. These shall be discussed here.

Non-response weights are based on the response propensity which is measured by the probability of response. Those whose characteristics lead to low response probability should have high weight values to represent more individuals from their category, since they are less likely to respond. In turn, individuals with characteristics that lead to high response probability should have low weight values to represent fewer individuals from their category, since they are more likely to respond. Thus, a non-response weight is basically the inverse of the response probability (propensity). This is why part of the literature in this area refers to non-response weighting as '*Inverse Probability Weighting (IPW)*'.

There are two ways to estimate the response probability for units in the sample in order to calculate the weights: weighting classes; and model-based methods.

Weighting classes

Weighting classes is a simple approach that involves dividing the sample into a number of non-overlapping sub-groups using a few auxiliary variables (also called weighting variables) that are known for both respondents and non-respondents (Kalton and Flores-Cervantes, 2003; Little, 1986; Brick, 2013 & Biemer and Christ, 2008). The resultant sub-groups are referred to as 'weighting classes'. The response probability for each weighting class is then calculated as the class response rate. The non-response weight that is assigned to a responding unit is simply the inverse of the response probability of the class to which the unit belongs. For example, for a given class 'c', if the number of units is denoted by n_c, and the number of responding units is denoted by m_c, the response probability is defined by m_c/n_c. Thus, the weight of a responding unit in class 'c' is

n_c/m_c. If there is homogeneity in terms of response propensity between all units in a weighting class (i.e. all units have the same response propensity), MAR holds, and non-response bias will be eliminated by using the weights. Therefore, an alternative term to weighting classes, which is sometimes used in the literature, is Response Homogeneity Groups (RHGs), see for example (Brick, 2013).

The disadvantage of this method is that classes are subjectively identified in one or two dimensions, by using one or two auxiliary variables. Also, classes with small number of respondents produce small response rates and, hence large weights. Larger values of weights may introduce large variances in estimates. Lynn (1996) suggests avoiding weighting classes with a response propensity that is less than one-fifth of the overall survey response rate.

Model-based methods

In a model-based method, usually a binary outcome regression model is used to estimate the response propensities for the sample units. This method was incorporated into the survey non-response problem by David *et al* (1983). It is an extension of the propensity score theory of Rosenbaum and Rubin (1983). Models used in this regard, often, are referred to as ‘Response Propensity Models (RPMs)’. With a suitable function, usually ‘logit’ or ‘probit’, the probability of response can be modelled (response=1; and non-response=0). The non-response weights can then be calculated, for responding units, as the inverse of the predicted values from the model. For example, if R_i denotes the outcome variable in a RPM that uses a logit function (i.e. logistic regression), R_i is an indicator with the following values (R_i= 1, if the i^{th} sample unit responds; and R_i= 0, otherwise).

Auxiliary variables (or weighting variables) that are available for both respondents and non-respondents, and are thought to be correlated with R_i can then be used to estimate the model.

For responding units, non-response weights are then computed as:

$$w_{NRi} = r_i^{-1} \tag{3}$$

Where w_{NRi} is non-response weight for unit i; and r_i^{-1} is the inverse of the predicted value of R_i.

Using a RPM to estimate the response probability for sample members may be more effective than applying the weighting classes approach (Grau, 2006). This is because a large mixture of dummy and continuous weighting variables can be used to fit a range of models, and therefore obtain more effective non-response adjustments. However, an important disadvantage is that the predicted response probabilities for some units may differ considerably. This may result in large weights variance. Large weights variance, in turn, will increase the variance of estimates. Nonetheless, the estimated response probabilities can be grouped into weighting classes, and weights can then be recalculated using either the mean predicted probability in the class or the observed response rate in the class.

Since we always assume that being selected in the sample is independent of responding to the survey, the weight that is usually used in the analysis is constructed as the product of the design weight and non-response weight. This way, every unit in the sample is adjusted using its chance of being selected in the sample and its tendency to respond to the survey simultaneously. The final analysis weight assigned to the i^{th} responding case is shown in equation (4).

$$w_i = D_i * w_{NRi} \tag{4}$$

Where w_i is the final analysis weight; D_i is the design weight; and w_{NRi} is the non-response weight.

Effect of weighting

Weighting is done to reduce bias in survey estimates. The underlying assumption for this is that characteristics of respondents in a weighting class (or with a given set of characteristics of the auxiliary variables that predict the probability of response) are similar to the unobserved characteristics of non-respondents in the same class with respect to the survey target variables (Lynn, 2005). When this assumption holds, weighting will then successfully reduce bias from estimates. However, there is a downside to weighting. That is variability in the weights will increase the variance of the survey estimates. Thus, while un-weighted estimates may be biased but more precise, weighted estimates are less biased but also less precise. This is an inevitable trade-off to be made in weighting. However, to limit the extent of the increase in variance, survey researchers sometimes restrict large weights to some arbitrary maximum value at which they can tolerate its corresponding increase in variance. This technique is referred to as 'trimming'.

Why use weighting?

Each of the post-survey adjustments has its advantages and disadvantages. They share similarities such as requiring auxiliary variables, but they differ in terms of the way they handle non-response. Weighting and calibration-based methods assign weights to respondents as compensation for those who are missing; whereas imputation methods attempt to estimate the missing values in the substantive variables. This raises the question as to whether weighting-based methods have advantages over imputation-based methods or vice versa. To give an insight into this issue, this section compares weighting with Multiple Imputation (MI) and the Selection Model approach (SMA).

Weighting relies on the MAR assumption. SMA works on the basis of NMAR. MI could be used under both MAR and NMAR, but the latter may require MI to correctly specify the model for missingness. MI views both unit and item non-response as a missing data problem. Consequently, it corrects simultaneously for unit and item non-response. Weighting, on the other hand, can only deal with unit non-response. Also, weighting ignores the association between the auxiliary variables and the outcome variable, which may lead to inefficiencies in the analysis (Plewis, 2011). Meanwhile MI and SMA take the association between the auxiliary variables and the outcome variable into account by establishing models that link the outcome variable and the missingness mechanism. Thus, the estimation of different substantive models may need the application of different MI and SMA models.

Weighting, however, is multipurpose. Once the weights are created they can be used in the estimation of different substantive models (i.e. the same set of weights is used every time). In addition, for secondary data users, who are concerned about the effect of non-response on their estimates, and who do not have the technical capabilities nor the necessary data to perform a procedure like MI or SMA, weighting may be a good option. It is relatively easy to use weights in most statistics software. The weights are in the form of a variable in the data set. Usually, users only notify the software that they would like to implement weighting in their analysis and simply identify the weighting variable. In return, the software carries out the necessary calculations and produces weighted survey estimates. Moreover, analysts are able to use many standard analysis techniques with weights.

Weighting in longitudinal surveys

It is common practise in longitudinal surveys that survey organisations prepare weights and include them in public use data files for use by analysts. Most of the household panel surveys implement a similar approach in terms of non-response weighting. To give an insight into this, in this section we describe and discuss this approach on the basis of two major surveys: the BHPS and HILDA; but more attention is paid to the BHPS since its data is used in this thesis.

The BHPS

Full details on the BHPS including the sample, survey instruments, fieldwork, measures and weighting procedures are well documented in Taylor (2010); Lynn (2006); and Uhrig (2008).

The BHPS was conducted in the period 1991 to 2008. It followed its sample members every year to conduct interviews. Its main purpose was to explore the dynamics of change experienced by the population in the UK. In addition, the BHPS was conducted so that secondary data users have micro-data sets available. These data sets can then be used to carry out a wide range of research across a range of social science disciplines, and for policy research. In general, the BHPS provides data in 9 main areas: labour markets, income, savings and wealth, household and family organization, housing, consumption, health, social and political values, education and training.

Eligibility to the BHPS was restricted to individuals who were residents in private households in the UK. Those who were not alive, not resident in the UK, or were in the UK but institutionalised (i.e. living in nursing homes, military bases or prison) were not eligible for the survey. Using the small user Postcode Address File (PAF) as a sampling frame, 8,217 addresses were drawn as original sample units. The frame included all countries in Great Britain except Northern Ireland. There were three stages of selection: using a systematic sampling technique, the first stage selected 250 postcode sectors from stratified listing of all sectors on the PAF as Primary Sampling Units (PSUs); in the second stage, fieldwork delivery points (equivalent to addresses) were selected from the resultant PSU from the previous stage using analogous systematic procedure; and a final selection stage was conducted by interviewers at the address level. During the selection of households, interviewers excluded non-residential addresses and institutions. A household in the BHPS was defined as “one person living alone or a group of people who either share living accommodation or share one meal a day and who have the address as their only or main residence”.

The first wave was conducted in 1991. Interviews were attempted with all household members who were aged 16 or over. This resulted in 10,248 individual interviews at wave

1. Subsequent to wave 1, the BHPS attempted following all sample members in wave 1 responding households and interviewing them as well as all new household members living with wave 1 sample members. Letters were sent to sample members, in subsequent waves, notifying them that the interviewer will call them within a week. Only adamant refusals were excluded from the fieldwork. Non-contacts were coded as such after six call attempts.

Weighting in the BHPS

The weighting in the BHPS is documented in volume A of the user manual (by Taylor, 2006).

To adjust for non-response, the BHPS calculates weights both at the individual and household levels. Our discussion will be limited to weights at the individual level, since the analyses here, and in most research, are done at the individual level. There are two types of weights: cross-sectional and longitudinal. Cross-sectional weights are available in every wave, but they are only suitable for cross-sectional analyses (single-wave analysis) in the corresponding waves. Longitudinal weights are available in every wave from wave 2 onwards. Longitudinal analyses that use data from a number of waves (any wave-combination of more than 1 wave) should use longitudinal weights from the last wave in the wave-combination in question. For example, to analyse data from wave 1, 10 and 18, or data from all waves up to wave 18, both scenarios should use the longitudinal weights at wave 18. In this section we describe the calculation of longitudinal weights.

At wave 1, there were two general types of weights: design weights and non-response weights. The design weights were derived to account for the different probabilities of selection due to the different stages of selecting the sample. These were calculated as the inverse of the probability of selection for every sample unit. However, our focus shall be on the creation of non-response weights. First, these were calculated at the household level using weighting classes. The variables used to identify the classes were region, socio-economic group (at the address level) and type of accommodation. In every class, the responding households were weighted by a factor that made their total number equal to the total number of responding and non-responding households in the class. A small number of cases within the responding households failed to respond at wave 1. However, information about these individuals were recorded during the household interview. To adjust for this, individual (within responding households) non-response weights were derived. A model-based method was used. By defining two outcomes: individual interview obtained=1; and individual interview was not obtained=0, a logit model was fitted. The variables used were age, gender, region, housing tenure, household size, marital status and employment status. For all responding cases, the weights were then defined as the inverse of the predicted probabilities from the model. These weights were then multiplied by the household non-response weight, and the resultant weights represent the individual non-response weights at wave 1. Note that BHPS does not release the design weights separately. The design weights were combined with the individual non-response weights from wave 1. Thus, the final analysis weight for a responding case at wave 1 is a product of the design weight and the individual non-response weight at wave 1 for that case. Final analysis weights in wave 1 represent the set of weights that is included in the BHPS wave 1 data file for analyses at the individual level.

In every subsequent wave, response was defined as responding in all waves up to and including the latest wave. In other words, both attrition and wave non-response were classified as absolute non-response. Non-respondents of unknown eligibility were treated as eligible non-respondents. The weights were then derived, every wave, (only) for those who responded at all waves up to the latest. Thus, the longitudinal weights at any wave are the product of subsequent weights accounting for losses between each adjacent pair of waves up to that point. Weighting in waves subsequent to wave 1, was done using weighting classes. A number of variables that were thought to be informative of non-response and of interest in the substantive analyses of BHPS data were used to form the classes. These variables include age, gender, race, employment status, income, education, region and tenure. At every wave, the method used variables from the previous wave. To make the process manageable, an automatic interaction detection programme (SPSS CHAID) was used to create the weighting classes. The weight for respondents in a given class was defined as the inverse of the response rate of that class.

Weighting in HILDA

In HILDA there are also two types of weights: cross-sectional and longitudinal. For HILDA, the weighting is documented in Watson and Fry (2002), Watson (2004) and Summerfield (2010).

At wave 1: First, the design weights were calculated as the inverse of the probabilities of selection of each household. Note that the design weight for a sample member in a household is the same as the design weight for that household, since the probability of selection of a person coincides with that of the household in which they reside. Second, these weights are adjusted for non-response at the household level at wave 1. These steps are exactly the same as those of the BHPS.

However, unlike the BHPS, which used weighting classes to create non-response adjustments, in HILDA, these were created based on a logistic regression model predicting the probability of response for each household (response=1; non-response=0). Data about both responding and non-responding households, that were used to estimate the model, came from: basic information about the selected households made by the interviewers; and from the 1996 census where descriptions about the neighbourhood to which the dwelling belongs are available (for the BHPS, the weighting variables came from the sampling frame). This information include: dwelling type, external conditions of the dwelling, security features of the dwelling, geographical location, density of area (population per square kilometre) and average household size. For responding households, the design weights were then multiplied by the inverse of the predicted probabilities from the model. Finally, and similar to the BHPS, within responding households, not everyone who was eligible for interview responded. However, information about these individuals was recorded during the household interview. This information was used in a logistic regression to predict the response probabilities for all persons. The variables used in the model include gender, age, marital status, labour force status, health condition, number of adults in the household, number of children in the household and housing tenure. For every responding person, the weight was calculated as the product of their inverse predicted probability from the model and their household weight (design and non-response). These are the initial persons weights that are available at wave 1 and suitable for cross-sectional analysis on data from wave 1.

The weighting steps from wave 2 onwards are also similar to those of the BHPS. The only difference is that HILDA uses a model-based method while the BHPS uses weighting classes.

At wave 2: The initial persons weights were adjusted for non-response at wave 2. The method used a logistic model to predict the probability of response at wave 2 given that the person responded at wave 1. Non-respondents of unknown eligibility were treated as eligible non-respondents. The variables used to estimate the model were from wave 1. These variables included: gender, age, age-squared, marital status, employment status, education, health status, number of children in the household dwelling type, tenure and region. For a responding person, the initial longitudinal weight was then multiplied by the inverse of their predicted probability from the model.

In every subsequent wave, the model is estimated using the same variables, from the previous wave, and the response probability in the current wave is predicted given that response is provided in the previous waves. For responding persons, the weights are then calculated as the product of their inverse predicted probabilities from the model in the current wave and their longitudinal weights from the previous wave. Thus, longitudinal weights in a wave '*w*' are only available for responding persons in all waves together up to wave '*w*'.

It is therefore obvious that calculating the weights in longitudinal surveys has a number of steps. First, the design weights are calculated in wave 1. The design weights are then adjusted by non-response weights in wave 1. The auxiliary variables used to create non-response weights in wave 1 depend on the information available about respondents and non-respondents. In some cases these are variables available from the sampling frame (as in the BHPS), and in other cases these are information from a national census or observed by the interviewer (as in HILDA). Once the survey has gone past wave 1, it becomes relatively straightforward to calculate the weights, but more importantly, weights in the current wave can be calculated using a large number of substantive variables (e.g. gender, age, education, etc…) from the previous wave. The response model at every wave predicts the probability of response in the current wave given that response was provided in the previous waves. The weights are then updated accumulatively, every wave, by the created weights in the current wave.

Both the BHPS and HILDA implement this approach. The difference is that BHPS uses weighting classes to predict the probability of response, whereas HILDA applies a model-based method. In any case, this approach, which is typical in household panel surveys, will be referred to, throughout this thesis, as the standard weighting approach (SWA).

Apart from offering a single set of longitudinal weights at every wave, the SWA has the following principles:

(a) Response is identified as responding in all waves up to the latest, and therefore weights are only provided for units responding in all waves up to the last one. In other words, those who skip responding in at least one wave are also identified as non-respondents and, thus, do not have weights (i.e. $weight_j$=0; where j denotes attriters, wave non-responders and complete non-participants).

(b) Those who do not respond because they have left the study population (e.g. deceased or out of scope) but are not known as such are implicitly assumed to be eligible non-respondents.

(c) After the first wave, the response probability is estimated using a mixture of common variables (from the previous wave) on which all sample members have measurements, and all sample members are used as one set in this estimation.

However, the complexity of household longitudinal surveys raises concerns (linked to the above points a, b and c) with respect to using the SWA. In this thesis, we view these issues as limitations in the SWA. In return, we design, discuss and evaluate three different alternative weighting approaches, corresponding to these issues, throughout the chapters of this thesis.

Concerns about the SWA

(1) Ignoring non-monotonic response patterns

Many household panel surveys implement a data collection policy that allows the possibility of wave-nonresponse followed by continuing participation. For a given respondent, this means that if he or she does not provide data in a particular wave (wave-nonresponse), the survey organisation can also attempt collecting data from them in a later wave. In this setting, the response pattern, in terms of the number of respondents, may differ across wave-combinations. Thus, a single set of weights at every wave – as in the SWA - may not be the best strategy to deal with this type of non-response. This is because this single set of weights, in a wave '*w'*, is designed by reference to responding in all waves up to wave '*w*'. Thus, it identifies those who skip responding in at least one wave between waves 1 and ‘*w*’ as non-respondents which may lead to excluding them, unnecessarily, from some analyses.

For example, in a 10-wave survey, if an analyst would like to analyse data only from respondents responding, say at waves 1, 5 and 10, regardless of their responding status in the rest of the waves, in the SWA the appropriate set of weights for this analysis is the longitudinal weights at wave 10. However, this set of weights will rule out respondents who are not present in all of the 10 waves by assigning a weight of zero to them even if they responded in waves 1, 5 and 10. Therefore, using the SWA may be inefficient in this case. A more suitable weighting strategy, in this example, would define response as responding to waves 1, 5 and 10; and non-response otherwise. However, this implicitly suggests that weights should be created for all possible combinations of waves, but his might be impractical if a large number of waves is conducted. Therefore, limited number of sets of weights may be a solution. Nevertheless, even the issue of identifying the specific wave-combinations for weighting needs appropriate investigation and practical evaluation.

(2) Treatment of unknown eligibility cases

In representative samples all sample members who provide data for the survey must be eligible (part of the population of interest) for the survey administration. Otherwise, if some respondents are ineligible, they should be excluded from the sample.

A common eligibility criteria for household longitudinal surveys is continuing to be alive and residing in the country where the survey is conducted. However, over time, eligibility status for some sample members may change. Sample members may die or move out-of-scope. The problem here is that, for some sample members, who were known to be eligible in earlier waves, eligibility status cannot be established in later waves if they cannot be contacted any longer. As a result, these will be classified as cases with unknown eligibility. Ideally, sample members whose eligibility is unknown but are actually ineligible should be excluded from weights creation (i.e. from the base of any weighting model). In the SWA, these are treated as eligible non-respondents. This may incorrectly increase the sizes of the weights in classes with more ineligible cases but are not known as such (recall that the weights are reciprocal of the probability of response). As a result, weighted estimates may be biased towards sample members from these classes.

However, identifying ineligible cases amongst cases of unknown eligibility is a challenging task. In waves subsequent to the initial non-contact, it is impossible, sometimes, to identify eligibility status at the case level for those whose eligibility is unknown. A promising alternative that may need a detailed investigation is to use population eligibility information to estimate eligibility rates for subgroups in the sample that contain unknown eligibility cases. The weights in these subgroups can then be adjusted based on the estimated eligibility rates.

(3) The choice of weighting variables and respondents for weighting

Often, longitudinal surveys target large populations (typically a living population in a country). Sampled units from such populations are not usually homogeneous with respect to survey participation. Some sub-groups are more cooperative than others i.e. the response propensity may be driven by different factors for different sub-groups. Thus, non-response predictors (weighting variables) in these sub-groups may differ from the general non-response predictors.

In general, some variables are believed to be better weighting variables than others. For example, age and employment status are known as good weighting variables because of their strong relationship with the response probability and most substantive analyses' outcome variables; while religious beliefs is not generally considered as a good weighting variable since it does not have a clear direct relationship with the response probability. Thus, using age and employment status together with other good predictors in weights creation generally yields a good set of non-response weights.

However, the same weighting variables may not be powerful in predicting the response probability in some sub-groups in the same sample. For instance, age and employment status might not effectively predict the probability of response in the sub-group of women aged 80 or over (because in this sub-group variation in age is minimal and all respondents are likely to be retired). In fact, variables such as religious beliefs may be a better predictor in this case. The point here is that using a common set of variables from the previous wave to create a single set of weights in the current wave – as in the SWA - does not necessarily result in a set of weights that can tackle non-response successfully in all sub-groups in the sample. Some sub-groups could use an alternative set of weights created from another set of variables. Some of these subgroups are important and are frequently used for analysis. This may be an alternative approach of weighting, but it needs practical evaluation.

Survey research to date has not yet investigated these issues (the concerns about the SWA) in the development of non-response weights in longitudinal surveys. In addition, it is not yet known, empirically, whether the raised concerns have considerable negative impact on the weights resulting from the SWA. Many on-going surveys apply the SWA. On the other hand, most data users do not have the statistical knowledge or the data to be able to construct any type of adjustment for the purpose of their analysis. Consequently, such data users, if they would like to adjust for non-response, rely on the weights provided by the survey organisations to reduce non-response error in their estimates. Thus, investigating these issues is useful from both the survey organisation and data users perspectives.

Moreover, the rapid rise in the improvement of survey designs (with the aid of current computer technology) which is faced by a rise in non-response rates, begs the question as to whether the SWA should also be improved to meet new challenges. However, at present, it is not known whether the SWA is able to deal with non-response error in all types of estimates. Also, other alternatives to weighting strategies are not being investigated extensively. Exploring other possibilities of weighting will benefit both surveys developing long-term weighting strategies and on-going surveys that need to improve their existing weighting. Thus, this thesis makes important and novel contributions to the development of non-response weighting for longitudinal surveys.

The research in this thesis sets out to investigate whether the SWA appropriately tackles non-response error in different types of estimates from different types of analyses with respect to the three raised concerns. The study aims to explore alternative weighting approaches (AWAs) to deal with non-monotonic response patterns, unknown eligibility whilst weighting and subgroup tailored-weighting. Also, the study evaluates the new approaches of weighting as opposed to the standard weighting approach. In the light of these objectives, the study seeks to answer the following questions:

1) Like any ordinary weighting approach, the SWA can deal adequately with the main aspects of the survey design. However, given that the SWA does not take into account a few important aspects of the survey that result from the longitudinal nature of the survey (non-monotonic response pattern, unknown eligibility and the choice of weighting variables and respondents), can the SWA deal with non-response error in all survey-based estimates?

2) If 'non-monotonic response pattern', 'unknown eligibility' and 'the choice of weighting variables and respondents' are taken into account to develop AWAs, will the AWAs have a different impact (in terms of magnitude and variance) on survey-based estimates compared to the SWA?

3) If the AWAs have a different impact on survey-based estimates as compared to the SWA, does this result in very different estimates (i.e. is the difference between the equivalent estimates resulting from the SWA and the AWAs significant)?

To achieve its objectives and answer its questions, this study is based on a specific analysis methodology. Aside from providing a rationalisation for the introduced approaches of weighting (AWAs), the analysis approach in this study is based on conducting empirical analysis on longitudinal data using the AWAs and the SWA, and comparing the results. It aims to provide practical examples for the evaluation of the AWAs based on longitudinal data from a major longitudinal survey. Namely, we use the

BHPS data for our evaluation. It is important to point out that the aim of our analysis in this thesis is not to compare the AWAs with each other. Rather, because each AWA is designed to address a particular limitation in the SWA, each AWA is compared in turn with the SWA to assess the effectiveness of the AWA in addressing the particular limitation.

In addition, one can focus on testing the differences between the new sets of weights resulting from the AWAs and the set of weights from the SWA (by conducting some forms of statistical tests). This strategy can enable one to report on differences across all sets of weights. However, implementing such analysis strategy will not provide information on the impact of the introduced approaches on estimates, and whether this impact is different than the impact of the SWA. This is because constructing survey estimates will be eliminated from the comparisons. With regard to the impact of the AWAs on estimates, practical evaluation is needed.

The remainder of this thesis is organised as follows:

Chapter 1 is concerned with the issue of non-monotonic response patterns. The SWA assumes monotonic response and therefore results in zero weights for any sample units that did not respond at every wave. This is suboptimal for any analysis that does not require data from every wave. The chapter therefore explores an alternative approach that involves designing sets of weights for wave-combinations that are more likely to be used for analysis. Two sets of weights are created based on the SWA and the AWA. Statistical analysis is conducted on the same sample using the two sets of weights separately, and results are compared to disclose differences between the two weighting approaches.

Chapter 2 focuses on investigating a method for estimating eligibility rates for subgroups in the sample where eligibility is unknown for some cases, using data external to the survey. The chapter then introduces the AWA which makes use of the estimated eligibility rates in the weights creation. Empirical analysis is conducted using the SWA and the AWA. Conclusions are drawn based upon comparisons between the results.

Chapter 3 is about the choice of weighting variables and the set of respondents used to create the weights. It explores another AWA (sub-group tailored-weighting), which recognises that the correlates of non-response could be different for different sub-groups. This approach is based on selecting a number of sub-groups from the sample and designing their weights by changing the weighting variables that are used in the SWA. Similar to the previous chapters, analysis is conducted using the SWA and the AWA. Evaluation of the AWA is based on comparing its results with results from the SWA.

CHAPTER 1

Weighting for Non-monotonic Response Pattern in Longitudinal Surveys

1.1 Introduction

The development of non-response weighting in longitudinal surveys requires paying attention to the complex aspects of the survey that emerge as a direct consequence of collecting measurements from the same units on multiple occasions. For an accurately weighted estimate, there should be a non-zero weight for all sample members who will contribute to the estimation process. Weighting that does not estimate weight values for, and therefore rules out, some of the respondents who should be used to construct the estimate in question, may not tackle non-response bias adequately. Moreover, it is an inefficient use of data in the estimation since some data are excluded even though they contain useful information.

After many waves are conducted, say *w* waves, estimates of the relevant longitudinal population at wave '*w*' can be constructed using data from all waves up to wave '*w*', or from a number of other possible combinations of waves that include wave '*w*', and that are subsets of the *w* waves. Different wave-combinations may be relevant to different analysis objectives. If, in every wave-combination, response is defined as responding in all waves in the combination, each sample member may be defined as either a respondent or non-respondent, depending on which wave-combination is under consideration. As a result, each wave-combination may result in a different set of respondents both in terms of number and composition. It is very likely that the number of responding units in larger combinations (wave-combinations that contain large number of waves) is smaller than the number of the responding units in wave-combinations with smaller number of waves.

Survey organisations, typically, offer one set of longitudinal weights at every wave for analysts who would like to adjust for non-response in their analysis. The weights are designed by reference to responding in all the conducted waves up to the current one. i.e. the designing and offering of the weights is done based on the principles of the standard weighting approach (SWA) that we set out in the introduction of this thesis. Thus, the weights in a given wave '*w*' adjust for the longitudinal non-response in all waves up to wave '*w*'. Consequently, weights are only available for those who responded in all waves up to '*w*'.

Such weights can be very useful in reducing bias in estimates that are constructed using data from a balanced panel from all waves up to '*w*'. However, weighted estimates relating to the longitudinal population at wave '*w*', which are constructed using data from a subset of any wave-combination that include wave '*w'*, will also use the longitudinal weights at wave '*w*' (the only offered longitudinal weights that are relevant to the longitudinal population at wave '*w*'). In this chapter, we investigate whether the latter is suboptimal. In other words, we investigate whether it is efficient to use the longitudinal weights at wave '*w*' to adjust for non-response in estimates based on data from subsets of waves that include wave '*w*'.

Our concern stems from the fact that weights from the SWA at wave '*w*' are only available for those who responded in all waves up to *w*. Whereas the responding sample in a subset of waves that include wave '*w'*, may contain sample members who did not necessarily respond in all waves up to *w*. As a result, with the SWA, these sample members will not be used to construct the weighted estimate in question, hence inefficient. Also, if these sample members (those who are part of the analysis sample but are not used to construct the estimate of interest as a result of using the SWA) have different values on the variables that are used to construct the estimate in question than the rest of the sample, the resultant estimate may be biased. These issues may arise if the response pattern during the course of the survey – as in many surveys- is non-monotonic. The SWA does not take this into account. Section 1.4 explains the missing link between the SWA and the non-monotonic pattern of response in detail.

With respect to a given wave-combination, weighting will be more appropriate if it estimates weight values for every respondent in this wave-combination irrespective of responding status in other wave-combinations. This implies that weighting should be done separately for all possible wave-combinations in terms of identifying response and the variables used to create the weights. However, creating subsets of weights for all possible combinations of waves may be impractical, especially after a large number of waves are conducted.

An alternative approach may be providing extra sets of weights for a limited number of combinations of waves. A challenge for the survey organisation then is to identify the best possible wave-combinations for the additional weighting.

In this chapter we introduce an alternative weighting approach (AWA) which creates sets of weights particularly for analyses restricted to specific wave-combinations. The AWA creates weights for wave-combinations that obtain data on the same theme. Such combinations are likely to be in demand from analysts as a base for analysis. We select one wave-combination of this nature from the BHPS. The selected combination contains data on wealth. For our evaluation, we create two sets of weights based on the SWA and the AWA respectively. We investigate whether the SWA is efficient in adjusting for non-response in data from the selected wave-combination. We also examine whether the AWA is more efficient in this case. The investigation is done by using the two sets of weights to analyse the same sample from the selected wave-combination separately, and compare the results.

In addition, the chapter reviews the types and sources of auxiliary variables in general, and specifies the type of variables that is used for our weighting; and it sets out the type of response propensity models that are used to create the weights in the thesis as these issues are fundamental when designing non-response weights.

1.2 Non-response weighting variables

As mentioned in the introduction of this thesis, post-survey adjustments rely on auxiliary variables in their treatment of non-response. For decades, the term 'auxiliary variables' was mainly used to describe variables that are not of analytical interest. In cross-sectional surveys, such variables are typically available from the sampling frame from which the sample was drawn. Also, auxiliary variables may be available from sources external to the survey, for example, from a national census. Therefore, auxiliary variables may be available for the full sample. In this sense, auxiliary variables are by definition not of substantive interest to the survey, as designing a survey to collect variables that already exist is unnecessary. However, when longitudinal surveys emerged, they provided the opportunity to use substantive variables that were collected in earlier waves as auxiliary variables to adjust for missingness in later waves. Auxiliary variables that are used in weighting adjustment in particular are sometimes referred to as 'weighting variables' (see for example Kreuter and Olson, 2011). In this research, regardless of the type of these auxiliary variables, we present them with the label 'weighting variables'.

The choice of weighting variables plays an important role in reducing non-response bias. In recent years, survey researchers have laid the foundation for principles to guide the selection of the best set of variables to adjust for non-response (Särndal and Lundstrom, 2005; Little and Vartivarian, 2003; 2005). A variable is said to be powerful in reducing non-response bias if: it shows evidence of explaining the response propensity, it is highly correlated with the survey main variables, and it identifies or comes close to identifying one of the important domains in the population. Little and Vartivarian (2005) demonstrated that if the association between the weighting variables and the variable of interest is low, the weighted mean will have increased variance without decreasing the bias even if the association between the weighting variables and the response propensity is high.

In sum, in order for non-response weights to be effective in reducing bias, the weighting variables have to be correlated with the substantive variable of interest and the response propensity. Additionally, to be able to create the weights, the weighting variables have to be observed for both respondents and non-respondents. However, even with fewer restrictions, the existence of a good set of weighting variables, in practice, may be rare. This is because, first, in practice, only a few variables are available for both respondents and non-respondents. Perhaps this is why, in recent years, survey researchers have extensively investigated alternative sources of variables that can be observed for all sample members, and advised survey organisations to move towards data collection modes that collect such variables. Second, even within the available variables, any given variable is likely to differ in the strength of its correlations with the substantive survey variables (Kreuter and Olson, 2011). Third, no single variable is likely to predict the response propensity and be correlated with all substantive variables simultaneously (Kreuter, Lemay and Casas-Cordero, 2007; Groves, Wagner and Peytcheva, 2007; Kreuter and Olson, 2011). This is why survey organisations should have plans to identify sources of potentially good variables and collect them at the data collection stage.

Weighting variables can be drawn from multiple sources. These sources could be internal or external to the survey. Depending on the type of variables and the source, the main categories of weighting variables are: (a) variables about the process in which the survey data were collected. This type of variables is referred to as 'paradata'; for example, what was the mode of data collection (phone, web, mail, or in person). (b) variables based on the interviewer's observations about some characteristics related to the household/individual (e.g. type of accommodation). (c) variables taken from the sampling frame, i.e. traditional auxiliary variables. These are usually available if the sample is taken from administrative records (e.g. levels of proficiency or education). (d) variables linked from another database. Sometimes the sampling frame does not provide much information about sample units, for example, if the sample frame is the postcode address file (Lynn, 1996). In this example, although the postcode itself does not provide information about sample members living at the selected address, it can be used to link geographical information from another database such as credit scores (Lynn, 1996). (e) substantive survey variables. In the case of a longitudinal survey, these variables could be available in previous waves.

While longitudinal surveys are fortunate with (e), some of the available literature focus on (a), (b), (c) and (d) (for example Plewis, 2011; Kreuter and Kohler, 2009; Lynn, 2003; Lynn, 1996). The advantage of paradata, interviewer observations, sampling frame variables and variables used to link information from another database is that they are cheap to observe if not completely free and can be available for every unit in the sample. For instance, variables related to the accommodation type, neighbourhood characteristics, time interviewer arrived to the house, and number of previous contact attempts do not require respondents to report them; instead, they can be observed by the interviewer.

Variables from (a), (b), (c) and (d) are successfully used in the literature to adjust for non-response. For example, using data from a number of surveys, Kreuter *et* al (2010) found that the inclusion of these variables in response propensity models that were used to derive non-response weights reduces the mean-square error (MSE) in measures of central tendency adjusted by the resultant weights. However, they found that very few of these variables are associated with the response propensity. In contrast, using Receiver Operating Characteristic (ROC) curves, Plewis (2011) assessed the impact of including these variables in the response propensity models. He found that their inclusion may improve the accuracy of the models (i.e. they are associated with response propensity), but they have little effect when adjusting for non-response. Also, Lynn (1996) showed how, in the Scottish School Leavers Survey (SSLS), information about the level of qualification gained at school which was available in the sampling frame was used to analyse the response rate in connection to making weighting adjustment. Lynn (1996) demonstrated the way in which the post code in the Health Survey for England (1994) was used to identify the area where the respondent lived as large urban/city centre, other urban/suburban or rural and then analysed the response rate accordingly.

In longitudinal surveys, variables from the categories (a) to (d) are usually used to adjust for non-response in the first wave. After the first wave, it is common to use key variables from previous waves (i.e. category (e) variables) to analyse and/or adjust for non-response in later waves. Most research has found variables such as gender, race, age, socioeconomic status, income and level of education to be good predictors of the response propensity and hence powerful weighting variables. For example, Watson (2004) states that from wave 2 onwards in HILDA, variables such as gender, age, marital status, labour force status, health condition in a current wave are used to create non-response weights in the next wave. Similarly, age, gender, race, employment status, income and education are used in the BHPS weighting after wave 1 (Taylor, 2006). Also, Siddiqui et al (1996) used proportional hazard regression in analysing the factors influencing dropout in longitudinal school-based smoking prevention studies; race, tobacco knowledge and academic performance were found to be significant factors. Kroh (2009) indicates that, in GSOEP, characteristics measured in 2007 (wave 23) such as gender, age, job status, income and savings were used to predict the probability of re-interviewing in 2008. Both Becketti et al (1988) and Fitzgerald et al (1998) showed that, excluding young respondents, attrition is positively associated with old age. Investigating attrition in the BHPS, Uhrig (2008) found that housing tenure, marital status, size of household, gender, race, region, mode of interview, employment, number of children in household, financial situation, education, health, income and social isolation are all associated with attrition.

In this thesis, the weights creation, in both SWA and AWA, is restricted to weighting variables from category (e). In every chapter, we use a model-based method to create the standard and alternative weights. Our response propensity models in a given wave include variables observed in previous waves. This enables us to take into account changes in respondents' characteristics over time which is likely to be reflected in the response propensities and hence in the weights.

1.3 Response propensity models for panel data weighting

Before introducing our AWA and discussing the SWA with regard to non-monotonic response pattern, it is important to present the response propensity models that will be used to derive the weights for both weighting approaches[4]. This section sets out the type of response propensity model that we use in the thesis in general. In each chapter, the model is modified depending on the research problem and the waves from which data are used in the analysis.

There are at least two methods to model the response propensity in order to derive weights for panel data: the first method estimates a marginal model at every wave. This model is defined based on the response status in the current wave conditional on response in the previous waves (note the response status in wave 1 is not conditioned on previous response). The overall response propensities are then estimated as the product of the predicted values from each of the wave-specific models, and the weights are set as the inverse of the overall response propensities. The second method creates wave 1 non-response weights separately, then uses wave 1 as a base. It then estimates one weighting model based on response in all the conducted waves conditional on responding at wave 1

[4] In this thesis we do not use the BHPS public-use weights. Instead, in each chapter, we design the weights that are relevant to the analysis for both the SWA and the AWA.

(response= responding at all waves; otherwise non-response). The resultant weights will then be multiplied by wave 1 weights.

Thus, both methods define response as responding in all waves together up to the last one. Both methods define the same set of respondents as responding (those who responded in all waves up to the latest). The differences lie in the form of the model and the set of weighting variables. The first approach models non-response as a series of steps, while the second treats non-response as a single process. The first method can use variables from the previous wave as covariates in each model, while the second method only uses variables from wave 1. Both methods create a set of weights that aims at reducing non-response bias in all estimates related to the longitudinal population at the last conducted wave.

The advantage of the first method is that it takes into account the fact that some respondents characteristics may change over time (time-variant variables), and therefore may have a different effect on the response propensity at different time points. However, this may be at the price of estimating larger weights (recall that the response propensity here is estimated as a product of the predicted values from all the wave specific models). The disadvantage of the second method is that it ignores the effect of time-variant variables on the response propensity, but may be more parsimonious compared to the first method as only one model is estimated.

To be able to take the effect of time-varying variables into account, in this thesis, we only apply the first approach. In other words, we only consider response propensity models that take into account changes in respondent’s characteristics over time.

For a simple illustration of this approach, consider a three-wave survey. Let R_1, R_2 and R_3 denote the response status in each of the three waves (R_t=1 if response is observed; R_t=0 otherwise; t=1,2,3). Let Z denotes a set of fixed characteristics related to sample members (individual/household characteristics or survey design variables) that are observed for all sample members at wave 1. Let X_t (t=1, 2, 3) denote a set of time variant variables that are collected in the three waves, with X_t observed if R_t=1. We can then estimate three logistic regressions:

R_1 on Z, using all sample members;

R_2 on Z and X_1, if R_1=1; and

R_3 on Z and X_2, if R_2= 1 and R_1=1.

If r_t are the estimated probabilities from model t (t=1, 2, 3), the weights for respondents in the three waves are: $w_1 = r_1^{-1}$; $w_2 = w_1 * r_2^{-1}$; and $w_3 = w_2 * r_3^{-1}$ respectively.

We apply the same approach on the BHPS data for the SWA, but we exclude sample members who become ineligible by the last wave in the waves-combination used for weighting. Also, we assume that sample members whose eligibility is unknown by the last wave in the waves-combination in question are eligible. For example, to construct the longitudinal weights at wave 10, we exclude sample members who become ineligible by wave 10 from the base of the weighting models used to create the weights as these cases are clearly not part of the longitudinal population that the weighting aims to represent. In addition, sample members whose eligibility is unknown by wave 10 will be assumed as eligible sample members.

As for the AWA, the method may be modified slightly depending on the issue under investigation. Also, note that one would want to use the resultant longitudinal weights in conjunction with the BHPS design weights to also account for the unequal probabilities of selection. Nonetheless, the BHPS does not release the design weights separately. These are released together (combined) with wave 1 non-response weights as explained in the introduction. Thus, we start modeling the response propensity from wave 2 onwards. This involves modeling the response in wave 2 conditional on responding in wave 1, modeling the response in wave 3 conditional on responding in waves1 and 2, and so forth. This way, the resultant longitudinal weights at wave '*w*', will compensate for the longitudinal non-response from wave 2 up to wave '*w*'. However, when these are multiplied with the BHPS wave 1 non-response/design weights, the resulting set of weights will adjust for non-response from wave 1 up to wave '*w*' as well as correcting for the unequal probabilities of selection.

Thus, our general response propensity model can be given by:

$$\textit{Logit}\ \Pr(R_{i,t}{=}1/\ C_{i,t-1}{=}1){=}\ f\ (\textstyle\sum_j \boldsymbol{\beta}_j Z_{ji}+\sum_k \boldsymbol{\beta}_k X_{ki,t-1}) \qquad (1.1)$$

Where *t* is the wave number for which the model is estimated (*t*=2, 3,…, 18); *i*= 1, 2, …, $n_{1,..,t-1}$, where $n_{1,..,t-1}$ is the number of respondents who responded at every wave from 1 to *t-1* and who are known or assumed as eligible by the time of wave T (T is the last wave in the waves-combination to which the analysis is restricted, $2 \le T \le 18$); $R_{i,t}$ is the response status at time (wave) *t* for respondent *i* ($R_{i,t}$=1 if response is observed at wave *t*; $R_{i,t}$=0 if response is not observed at wave *t*); $C_{i,t-1}$=1 if $R_{i,b}$=1 for all values of b from 1 to *t-1* (i.e. $C_{i,t-1}$=1 indicates that the model in wave *t* is conditioned on response in all of the previous waves); Z_{ji} is the set of time invariant variables for respondent i; $X_{ki,t-1}$ is the set of time variant variables for respondent *i* which are measured in wave *t-1*.

The model in (1.1) represents the base of our response propensity models in this thesis. Depending on the issue and the waves used in each chapter the model can be modified, for both the SWA and the AWA, as will be discussed in the relevant chapters.

1.4 Non-monotonic response pattern and the SWA: what is the missing link?

In longitudinal surveys, survey organisations may implement one of many data collection policies in terms of identifying the set of sample members who will be contacted for interviews at every wave. The major data collection policies are: attempt collecting data from sample members at every wave regardless of whether they participated in a previous wave or not; attempt collecting data only from wave 1 responding sample members or attempt collecting data only from sample members responding in the previous wave.

A typical scenario in household panel surveys, including the BHPS, is the collection of data at every wave, if possible, regardless of participation in previous waves. Such policy of data collection is advantageous because it provides an opportunity to potentially collect data, at some point during the course of the survey, from sample members who are hard to contact or reluctant to participate. However, it is likely to result in a non-monotonic response pattern in which wave non-response can take place unconditionally at any wave(s) during the course of the survey. As a result of a non-monotonic pattern of response, the number of responding sample members will differ across different combinations of waves. For example, table 1.1 shows the effect of the BHPS non-monotonic response pattern in the first three waves. The table displays the number of responding sample members in all possible combinations of waves. With just three waves, there are seven possible combinations of waves in which a sample member might respond. However, each of these combinations has a different responding sample size. In general, responding in a larger number of waves is associated with a smaller sample size and vice versa. For instance, the number of those who responded in wave 1 and 3 is 8, 419. Whereas 8,170 sample members responded in all of the first three waves.

To understand the limitation of the SWA with regard to non-monotonic response pattern, four groups of sample members can be distinguished: a) those who have measurements in all waves (a balanced panel); b) those who only have measurements in wave 1, or in wave 1 and the next consecutive wave(s), but they have no measurements in all of the waves (atritters); c) those who have measurements in one or more waves that are not necessarily consecutive, but they do not have measurements in all waves of the survey (wave non-responders); and d) those who did not provide measurements at all during the course of the survey (outright non-responders).

The SWA in a wave '*w*' produces longitudinal weights to adjust for the missingness of sample members from group (b) to (d) in analyses that estimate parameters of the population at time (wave) '*w*'. However, depending on the wave-combination that is used in the analysis, the SWA may be suboptimal with regard to adjusting for sample members in group (c). For example, consider the first three waves of the BHPS in table 1.1. The standard longitudinal weights in wave 3 are available for those who responded in wave 1, 2 and 3 (i.e. for 8,170 respondents). Those who did not respond in at least one of these waves their weights values are zeros. The weights are designed to adjust for non-response in estimates relating to the population at wave 3. Thus, for analysts who are constructing estimates by using respondents in waves 1, 2 and 3 (8,170 respondents), the weights are appropriate.

However, for analysts who would like to construct weighted estimates relating to the longitudinal population at wave 3 (and therefore will use the longitudinal weights at wave 3), but would only like to use respondents in wave 1 and 3 (8,419 respondents) the weights are suboptimal. This is because 249 respondents (who skipped responding in wave 2) from the sample in this case will be assigned a zero-value weight. As a result, the weighting will cause an unnecessary loss of some respondents.

If the weights are still used in the latter case, the sample size will be reduced to 8,170 respondents instead of 8,419 respondents. This may potentially result in a different estimate because the two sets of respondents are different both in terms of size and composition. Weighted analyses that only use respondents from wave 1 and 3, will benefit from a weighting approach that estimates weights values for all of the 8,419 respondents who are present in wave 1 and 3. i.e. a weighting approach that identifies response as responding in wave 1 and 3 regardless of responding in other waves.

In short, as a result of non-monotonic response pattern, different wave-combinations may result in different sample sizes. The SWA ignores this effect of non-monotonic response pattern and produces a single set of weights at every wave '*w*' for those who responded in all waves up to '*w*'. This set of weights is useful in analyses using a balanced panel from all waves up to *w*. However, it may be suboptimal in analyses that only use data from a subset of the '*w*' waves because some cases in the sample, in this case, will be given zero-value weights.

Table 1.1 Number of responding units in different wave-combinations in the first 3 waves of the BHPS

	Wave 1	Wave 2	Wave 3	Number of respondents (%)
1				10,248 (89.18)
2				9,845 (85.68)
3				9,600 (83.44)
4				8,970 (78.06)
5				8,736 (76.02)
6				8,419 (73.27)
7				8,170 (71.10)

* The shaded areas indicate the wave-combination of which the number of responding sample members is given.% were calculated out of the number of the selected sample.

1.5 The alternative weighting approach

In theory, the way to account for the effect of non-monotonic response pattern is to design a subset of non-response weights for every possible combination of waves. However, providing weights for all possible combinations of waves might not be achievable in practice after several waves are conducted. After k waves, there is a (2^k-1) possible combination of waves to provide weights for. For instance, with just 10 waves, there will be 1,023 possible combinations of waves that weights can be created for; and the number increases rapidly when more waves are added. In addition, in practice, it is unlikely that every possible combination of waves will be used separately for substantial weighted analysis. Some wave-combinations may not be of substantive analytical interest.

Thus, alternatively, a limited number of subsets of weights may be produced for a limited number of wave-combinations. Our alternative weighting approach in this chapter rests on this strategy.

Although the number of the limited subsets of weights may be decided subjectively, such strategy could be more useful than a single-set of weights strategy, and more practical than designing all possible subsets of weights. Figure 1.1 illustrates this in a simple diagram. The figure shows that there are at least three possible strategies of weighting with respect to the number of weights sets: single set of weights (as in the SWA), all possible sets of weights and limited number of sets of weights. The limitations of the first and the second strategies are that they could be suboptimal and impractical respectively. The advantage of the third strategy is that it is practical, but may also be considered as 'optimal'. The word 'optimal' here does not reflect maximum statistical precision since not all possible sets of weights are created. It, rather, indicates that a limited number of sets of weights may be a good compromise.

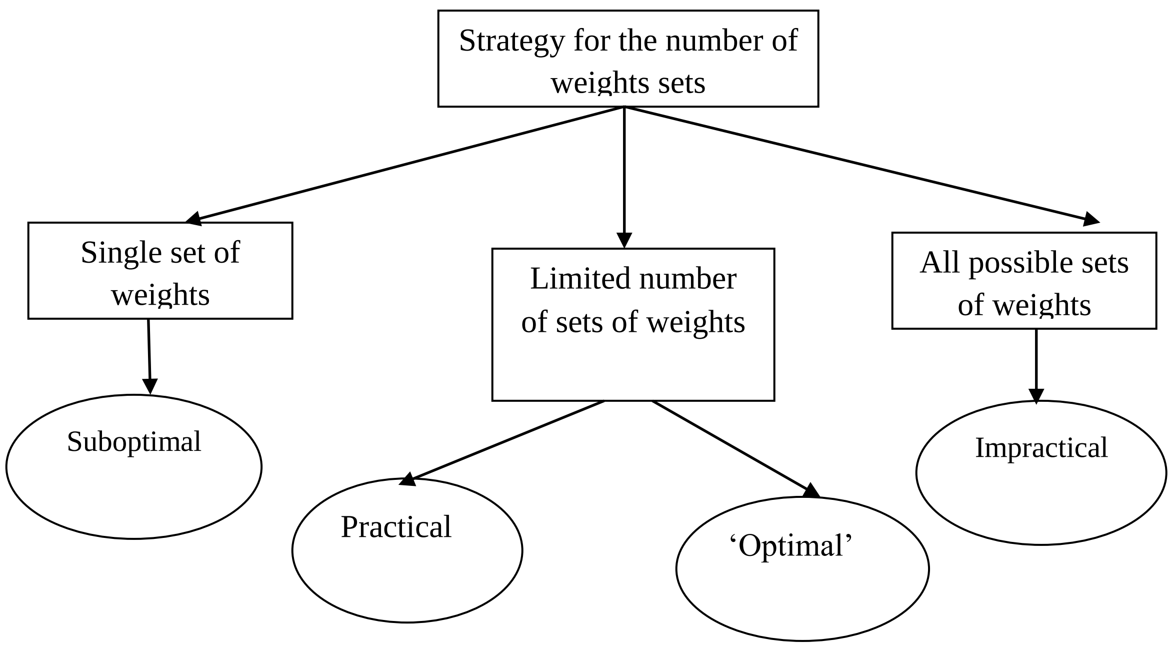

Figure 1.1. Three types of weighting strategy with respect to the number of weights sets.

Even if a limited number of sets of weights is considered, it is a challenging task to identify which specific wave-combinations should be selected in order to design their weights. There is very little literature in this area. In fact, the only effort that we are aware of is by Lynn and Kaminska (2010), suggesting criteria for selecting limited number of wave-combinations for weights creation. According to Lynn and Kaminska (2010), the following criteria should be considered when choosing combinations of waves:

Survey Design: If a certain combination of waves is not available in the survey by design, it can be excluded. They give an example by stating "*if a survey has a rule not to attempt data collection from any unit that has been non-respondent in three consecutive waves, then all combinations involving a respondent wave following three or more non-respondent waves can be dropped*".

Analytic Use: Weights should be produced for combinations that are more likely to be wanted by analysts.

Level of Non-response: If the samples responding to two wave-combinations differ only by a few cases, it is unlikely that weights derived for one combination will make much difference to analyses for the other combination. Thus, one subset of weights could be derived for both wave-combinations.

Correlates of Non-response: If the non-response process is very similar amongst a consecutive set of waves in terms of the covariates predictive of non-response, weights designed for a subset of waves from these consecutive waves might be similar to weights designed for the whole set. This may not include wave 2 and 3 as the attrition in these is believed to be distinctive. For example, there is evidence in the literature (e.g. Watson and Wooden, 2009) that correlates of non-response at wave 2 and 3 are distinctive, but thereafter are similar across subsequent waves. Thus, weights derived for response at wave 1, 2, and 3 might be quite different from weights derived specifically for waves 1 and 2, whereas weights derived for waves, say, 8, 9 and 10 might be quite similar to weights derived for waves 8 and 9.

Impact on Estimates: This could be used to judge the consideration of the criteria used to identify the combination of waves for which weights should be produced. For example, subsets of weights that produce the same estimate as others could be dropped.

For our AWA, the choice of wave-combinations for weighting is guided by the fact that some substantive estimates of the population can only be generated from data in specific wave-combinations. The idea of the AWA therefore, is based on providing extra sets of weights particularly for such combinations of waves. Thus, if the estimates in question need to be adjusted because of non-response, the AWA will have an advantage over the SWA. For example, a common feature of longitudinal surveys is a frequently asked module of questions where certain waves are conducted to obtain information about specific topic(s). For instance, wave 8, 13 and 18 in the BHPS provide data on neighbourhood, expectations of relationships and marriage in future. Such a wave-combination is likely to be used separately to provide estimates about the social phenomena that are measured in its waves. Unlike the SWA, the methodology of the AWA is to design weights for this wave-combination by defining response as responding in waves 8, 13 and 18 regardless of response status in the other waves. As a result, non-zero weights will be available for all responding sample members in this wave-combination. Thus, weighted analysis based on waves 8, 13 and 18 may use the alternative set of weights to avoid the potential loss of respondents that is associated with the standard weights at wave 18.

1.6 Methodology

To investigate the issue of non-monotonic response pattern in relation to the SWA, and to evaluate our introduced AWA, we selected a combination of waves from the BHPS that consists of waves 1, 5, 10 and 15. Although all BHPS waves, generally, provide data to be used for analysis in many of the social science disciplines, some waves are designed to cover certain topics extensively. Our chosen wave-combination collects data about wealth, assets and debt. Data from such subsets of waves might be used separately in studies of wealth dynamics and associated phenomena. However, the BHPS does not provide subsets of weights that are designed especially for the analysis of this combination of waves. The weights that are available for the analysis of data from this combination are the longitudinal weights at wave 15 (recall that these are designed by reference to responding in all the 15 waves, and are available for a balanced panel from wave 1 to 15, i.e. SWA).

For our investigation in this chapter, we designed a set of longitudinal weights at wave 15 using the SWA (response is defined as responding in all waves from 1 to 15). Also, we designed an alternative set of weights by identifying response as responding in waves 1, 5, 10 and 15 regardless of responding in the other waves. The latter is presented as the AWA in this chapter. Both the SWA and AWA modeled the response propensity using the model in equation (1.1). Details on these models and weights construction are provided in the next section.

Our aim is to use our standard and alternative weights in the analysis of savings and debts based on the responding sample in the wave-combination under investigation (a balanced panel from waves 1, 5, 10 and 15).We compare estimates resulting from the standard and alternative weights to provide evidence for potential differences between the SWA and AWA. These details are in section 1.7.

Potential differences between the SWA and AWA are due to differences in the sample members that each approach defines as responding.

To explore differences between the responding units at waves 1, 5, 10 and 15 and in all waves up to wave 15, table 1.2 shows the number of respondents in these two combinations of waves. The table also presents (in brackets) the proportions of wave 1 respondents for the two combinations. The number of respondents in waves 1, 5, 10 and 15 is 5,132 (50.08% of those who responded at wave 1). This is 4.7% higher than the number of respondents in all the 15 waves 4,654 (45.41% of those who responded at wave 1). This difference is caused by 478 respondents who participated in waves 1, 5, 10 and 15 but failed to respond in at least one other wave between 1 and 15. These results indicate that the standard weights at wave 15 are only available for 4,654 respondents who are identified as respondents by the SWA. Consequently, when the set of standard weights is used in the analysis of savings and debt which only uses the responding sample in waves 1, 5, 10 and 15 (5,132 respondents) it will rule out 478 respondents from the analysis. This is despite the fact that these sample members are actually respondents in these waves. This loss represents 9.31% of the balanced panel in waves 1, 5, 10 and 15. In contrast, the AWA produces weights for 5,132 respondents taking into account the 478 cases missing in the AWA. This is because the weighting model, in the AWA identifies these 478 cases as respondents. Therefore, in our analysis, it seems reasonable to expect more precise estimates from using alternative weights than from using standard weights.

As for the bias, we do not have specific expectation for this as it cannot be assumed that the AWA will result in less biased estimates than the SWA. Even though the weighting model in the AWA identifies a different set of sample members as respondents by avoiding the loss of 478 cases, we cannot easily assume that its weights result in less biased estimates compared to weights from the SWA. The bias may be reduced if the additional 478 respondents are similar, in their savings and debt characteristics, to non-respondents in wave 1, 5, 10 and 15. With data observed only for respondents, this assumption cannot be supported. Thus, in our comparison between estimates resulting from the two sets of weights, the analysis will focus on levels of precision rather than bias reduction.

Table 1.2 Number of respondents and non-respondents at waves 1, 5, 10 and 15 and in all waves up to 15.

	Respondents	Non-respondents	Total
Waves 1, 5, 10 and 15	5,132 (50.08%)	5,116 (49.92%)	10,248
All waves up to 15	4, 654 (45.41%)	5,594(54.59%)	10,248
Difference	478		

* The total indicates the number of those who responded at wave 1. The percentages in brackets indicate proportions of wave 1 respondents.

1.6.1 Construction of standard and alternative weights

This section describes how the response propensities were modelled in the SWA and the AWA, and the derivation of the weights associated with each approach. Based on the principles of the SWA, standard longitudinal weights at wave 15 were designed by defining response as responding in all of the 15 waves. Whereas in the AWA response was defined as responding in waves 1, 5, 10 and 15 regardless of responding status in the other waves.

To apply the model in equation (1.1) in this context: for the SWA, the same model in equation (1.1) can be estimated by varying *t* from 2 to15 (i.e. estimate 14 models, one model at each wave starting from wave 2 up to wave 15); and then calculate the weights as the product of the inverse predicted probabilities from all of the models[5]. Note that according to equation (1.1) the model at wave *t*, models response at wave *t* conditional on responding in all of the previous waves, but it uses variables from wave *t*-1. Meanwhile for the AWA, we are only interested in response status in waves 1, 5, 10 and 15. Thus, *t* can only range between 5, 10 and 15 (i.e. only 3 models can be estimated). In other words, we will model the response in wave 5 conditional on responding in wave 1 (and use variables from wave 1), model the response in wave 10 conditional on responding in waves 1 and 5 (and use variables from wave 5), and then model the response in wave 15 conditional on responding in waves 1, 5 and 10 (and use variables from wave 10), and calculate the alternative weights as the product of the inverse predicted probabilities from the three models. If this strategy is followed, aside from being different in terms of the way they define response, the two weighting approaches are also likely to differ in terms of the variables used to create the weights.

However, to investigate differences between the resultant sets of weights, ideally, we would like to use the same set of covariates to create both sets of weights so that any

[5] We did not use the standard longitudinal weights at wave 15 which are provided by the BHPS because these weights were created using a different approach (weighting classes) and a different set of covariates. To be able to examine potential differences between the SWA and AWA, weights resulting from both approaches should be created based on the same method and the same set of covariates so that differences can be said to be due to the different sets sample members that each approach defines as responding.

potential difference between the two sets of weights can be said to be due to differences between their weighting approaches in terms of how response is defined.

Thus, for the SWA, we applied the model in equation (1.1), but inclusion of the time variant variables in each model was restricted to variables from wave 1, 5, 10 and 15. That is: models from wave 2 to 5 used variables from wave 1; models from wave 6 to 10 used variables from wave 5; and models from wave 11 to 15 used variables from wave 10.

For those who responded in all of the 15 waves, the longitudinal weights at wave 15 were calculated as the product of the inversed predicted probabilities from all models, and wave 1 non-response/design weights (provided by the BHPS) as shown in equation (1.2).

$$SW_i = D_i * \prod_{t=2}^{15} r_{ti}^{-1} \quad (1.2)$$

Where SW_i is the standard longitudinal weight at wave 15 for respondent i; r_{ti} is the predicted probability for respondent i from wave t model (t= 2, 3,…, 15); i= 1,..., $n_{1,..,15}$ (where $n_{1,..,15}$ is the number of sample members who responded at every wave from 1 to 15); and D_i is wave 1 non-response/design weight for respondent i.

Those who did not respond in all of the 15 waves had their weights set to 0.

Turning to the AWA, the response propensity was modelled as follows:

In wave 5, we modelled the response propensity conditional on responding in wave 1 using fixed characteristics and time variant variables from wave 1 as shown in equation (1.3).

$$Logit\ \Pr(R_{i,5}=1 / R_{i,1}=1) = f\left(\sum_j \boldsymbol{\beta}_j Z_{ji} + \sum_k \boldsymbol{\beta}_k X_{ki,1}\right) \quad (1.3)$$

Where $R_{i,5}$ and $R_{i,1}$ are the response statuses in wave 1 and 5 respectively (R_i= 1 if response is observed; R_i=0 if non-response, for both waves); i=1, 2,…n_1, where n_1 is the number of respondents in wave 1 who are known or assumed as eligible by the time of wave 15; Z_{ji} is the set of time invariant variables for respondent *i*; and $X_{ki,1}$ is the set of time variant variables measured in wave 1 for respondent *i*.

In waves 10 and 15, modelling the response propensity is given by equation (1.4).

$$Logit\ \Pr(R_{i,t}=1/\ C_{i,t-1}=1)=f\ (\textstyle\sum_j \boldsymbol{\beta}_j Z_{ji}+\sum_k \boldsymbol{\beta}_k X_{ki,t-5}) \tag{1.4}$$

Where *t* is the wave that we estimate the model for (*t*=10, 15); *i*= 1, 2,…, $n_{1,..,t-5}$, where $n_{1,..,t-5}$ is the number of respondents who responded at every wave (in the waves-combination in question) from 1 to *t*-5 and who are known or assumed as eligible by the time of wave 15; $R_{i,t}$ is the response status for respondent *i* at wave *t* (*t*=10, 15; and $R_{i,t}$=1 if response is observed; $R_{i,t}$=0 if response is not observed); $C_{i,t-1}$=1 if $R_{i,b}$=1 for all values of *b* (in the waves-combination in question) from 1 to *t-5* (i.e. the model in wave 10 is conditioned on $R_{i,1}$=$R_{i,5}$=1, whereas the model in wave 15 is conditioned on $R_{i,1}$=$R_{i,5}$=$R_{i,10}$=1); Z_{ji} is the set of time invariant variables for respondent *i*; and $X_{ki,t-5}$ is the set of time variant variables for respondent *i* (measured at wave *t*-5).

The longitudinal weights at wave 15 for those who responded in waves 1, 5, 10 and 15 were calculated as the product of the inversed predicted probabilities from the three models as shown in equation (1.5).

$$AW_i=D_i * r_{5i}^{-1} * r_{10i}^{-1} * r_{15i}^{-1} \tag{1.5}$$

Where AW_i is the alternative longitudinal weight at wave 15 for respondent *i*; r_{ti} is the predicted probability for respondent *i* from wave *t* model (*t*= 5, 10, 15);); *i*= 1,..., $n_{1,5,10,15}$ (where $n_{1,5,10,15}$ is the number of sample members who responded at wave 1, 5, 10 and 15) and D_i is wave 1 non-response/design weight for respondent *i*.

Those who did not respond in all of waves 1, 5, 10 and 15 had their weights set to 0.

Although 10,248 sample members responded at wave 1, both the SWA and the AWA restrict the analysis to respondents known or assumed to be part of the longitudinal population at wave 15 (8,961 cases). Our choice for the weighting variables (in both SWA and AWA) was guided by the variables used in the BHPS weighting, and the availability of the variables across the waves, particularly in waves 1, 5 and 10 as some variables are not measured at every wave. The variables used are: gender, race, age, age-squared, health status, tenure, presence of children in the household, education, type of household, employment status, type of house, number in full-time employment in household and region. The results from the best models, for both the SWA and AWA, are displayed in table 1.3 and 1.4 respectively. Both tables present odds ratios. We consider coefficients to be significant if the relevant $p\text{-value} < 0.05$.

The results of modelling the response propensity in the SWA and the AWA seem similar in general. Also, the results show that the most of explanatory variables used in the analysis are correlated with response propensity. These results are discussed, for both weighting approaches, in what follows:

Gender: in most longitudinal surveys, females are more likely to respond than males (e.g. Hawkes and Plewis, 2006). Here we also find this to be the case in both the SWA and AWA.

Ethnicity: the literature usually indicates that ethnic minority groups show lower tendency to response in comparison with the majority groups (Gray *et al.*, 1996; Lepkowski and Couper, 2002). Ethnicity here is included in the models as a dichotomous variable specifying if respondent is white or non-white[6]. This was only significant in the models at wave 5 and 10 of the AWA indicating that white respondents are more likely to respond compared to other ethnic groups.

Age: age is known to be a good predictor of the response outcome. Excluding the oldest respondents, higher response is likely among older respondents than among their younger counterparts (e.g. Uhrig, 2008). In our models, apart from age, we also included age-squared to capture variability in the oldest age groups. In both SWA and AWA, age and age-squared are significant in most models. However, while the increase in age is positively associated with the response, increase in age-squared is associated with non-response indicating that respondents in the oldest age groups are less likely to respond.

Health condition: this has five categories (excellent, good, fair, poor and very poor). In this analysis, these categories were reorganised. The first three categories were combined into one category (good health) and the last two were combined into another category (bad health). The former is the omitted category. The results show that those with a bad health status are less likely to respond than those with a good health status. This is more so (and more significant) in latter waves than in earlier waves as the effect of poor health on response is clearer. This result is in line with most of the literature in this area (Nicoletti and Buck, 2004; Lepkowski and Couper, 2002; Becketti *et al*., 1988).

Housing tenure: this has two categories in our models: homeowners and non-home owners. Most research found that homeowners are more likely to respond compared to those who do not own their homes (Becketti *et al*., 1988; Fitzgerald *et al.*, 1998;

[6] The sample size for ethnic minority groups in the British Household Panel Survey sample are too small to allow for valid analysis of different ethnic groups.

Lepkowski and Couper, 2002; Watson, 2003; Nicoletti and Peracchi, 2005). Our explanation for this is that: maybe those who own their homes are less likely to move house and therefore are easier to establish contact with in successive attempts. Although housing tenure is only significant in a few models, its results confirm the general findings in the literature for both SWA and AWA.

Education: in our models education is represented by a dummy variable indicating whether the respondent has a General Certificate of Education (GCE) level A-C (or equivalent) or above. Most results in tables 1.3 and 1.4 show that those who have a GCE or any higher qualification are more inclined to respond. This result is confirmed by a number of non-response studies (e.g. Watson, 2003; Gray *et al.*, 1996; Lepkowski and Couper, 2002).

Employment status: this is known to be problematic when trying to understand its relationship with response propensity (Watson and Wooden, 2004). Some studies found that the response propensity is high among unemployed sample members (e.g. Watson, 2003; Nicoletti and Peracchi, 2005). This can be explained by the fact that unemployed sample members spend more time at home, and hence are more likely to be contacted. Other studies showed that there is a higher tendency to response among employed respondents (Gray *et al.*, 1996; Lepkowski and Couper, 2002). The explanation for this may encompass the fact that employed sample members are more likely to be geographically immobile and hence easier to follow over time. Our results here are also mixed. When employment status is significant, most models show that employed individuals are less likely to respond than unemployed ones. However, the model in wave

14 of the SWA shows that employed individuals are more likely to respond than those who are unemployed.

Presence of others during interview: if others are not present during the first interview, respondents tend to discontinue participating in the survey. This is likely because the presence of others is correlated with the household size and, therefore, with the possibility of making contact in subsequent waves (Uhrig, 2008). This is confirmed by a number of models in tables 1.3 and 1.4.

Type of household: we distinguished between two types of households: single-person household and multi-person household (Reference category). Non-response literature indicates that response is more likely among multi-person household than amongst single-person household (Brehm, 1993; Groves and Couper, 1998; Groves *et al*, 2002). This is partly because some individuals who live alone are less inclined to interact with others; and partly because establishing contact is less likely with single-person households than with multi-person households. When this variable is significant in our models, it indicates that single-person households are less likely to respond than multi-person households.

The presence of children in household: this variable was included in the models under the assumption that the presence of children in a household increases the chance of contacting the household. Households with children are easier to locate and establish contact with since the presence of children is associated with residential stability and community integration activities such as taking the kids to the nursery or school. This variable was found significant in predicting response in a few waves as shown in tables 1.3 and 1.4.

The results support our assumption that households with children are more likely to respond than households without children.

Type of accommodation: the type of accommodation that sample members reside at can influence likelihood of response. This is mainly because of potential access impediments attached to particular types of accommodation which may negatively affect the success of contact attempts (Groves and Couper, 1998). Prior research (Uhrig, 2008) found that living in blocks of flats where access to a number of apartments is through one entrance is more associated with non-response than living in houses that have their own entrance. In our analysis, 'type of accommodation' has three categories: living in a house (reference category) including detached, semi-detached terraced house; living in a flat; and living in other type of accommodation. The results in tables 1.3 and 1.4 show that this variable is only significant in a few models. However, our results are in line with the general findings of non-response literature.

Number of employed individuals in household: on the one hand, the number of household members can be positively associated with the probability of making contact with the household. On the other hand, the chance of non-contact is higher if more household members are in full-time employment. Our results here show that, in some models, households with a number of persons in employment are less likely to respond than households with no person in employment (e.g. the models in waves 11 and 10 of the SWA and AWA respectively). In other models, it is the exact opposite (e.g. the models in waves 13 and 15 of the SWA and AWA respectively).

Region: region is included in the models as a number of dummies with London as the omitted category. In models where region is significant, the results are similar in which the majority of the areas are more likely to respond than London. This applies for both the SWA and AWA. This result is consistent with the findings of Uhrig (2008) in his analysis of BHPS attrition. Uhrig reported that the South-East, South-West, East Anglia and the North-East are more likely to respond than London.

Table 1.3 Response propensity models based on the SWA (wave 2 to 15): modelling response in wave *t* conditional on responding in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9	Wave 10	Wave 11	Wave 12	Wave 13	Wave 14	Wave 15
Female	1.15*	1.36***	1.09	1.25*	1.06	1.14	1.37**	1.18	1.11	0.96	1.20	1.21	1.31*	1.51***
White	1.02	1.57**	1.47*	1.61**	1.21	1.49**	1.53*	1.09	1.64**	1.56*	0.99	1.63*	0.84	1.55*
Age	1.07***	1.08***	1.07***	1.08***	1.06**	1.04*	1.11**	1.00	1.00	1.09***	1.12***	1.10***	1.10**	1.12**
Age-squared	0.99***	0.99***	0.99***	0.99***	0.99**	0.99*	0.99**	1.00	1.00	0.99***	0.99***	0.99***	0.99**	0.99**
Bad health	1.12	1.09	1.11	0.82	0.85	0.66*	0.93	0.88	0.89	0.72*	0.66*	0.56***	0.88	0.51***
Home owner	1.19	0.94	1.11	0.97	0.84	1.07	0.98	1.22*	0.86	1.26	0.97	1.17	1.44*	0.98
Has GCE qualification or above	0.92	1.21*	1.25*	0.89	0.87	1.08	1.25*	1.05	1.15*	0.89	1.46**	1.07	1.55**	0.90
Employed	0.85*	0.89	0.83*	0.196	1.06	1.12	0.88	0.90	0.92	1.05	0.73*	0.72*	1.33*	0.82
Others present in interview	1.11	1.05	0.91	0.95	1.42**	0.90	0.87	0.89	0.90	0.91	1.36*	1.24*	1.53**	0.86
Single-person household	1.09	0.64*	1.06	1.09	0.76***	0.90	0.84	1.06	1.04	0.62**	0.96	0.80*	0.92	0.91
Household with children	1.49***	1.07	1.25*	1.04	1.03	0.82	1.01	1.02	0.98	1.05	1.71**	0.97	0.89	0.85
Living in a flat	0.78	0.77	1.22	1.21	0.74	1.29	1.15	0.87	0.82	0.91	0.78	0.89	0.61*	0.92
Living in other type of house	1.27	1.40	1.60	1.51	0.76	0.64*	1.15	0.76	0.70*	1.16	0.72	0.96	0.83	1.62
1 or 2 persons in employment	1.21	1.30	1.07	1.04	0.79	0.77	0.91	1.36	1.01	0.81	1.02	0.86*	1.11	1.24
3 + persons in employment	0.97	0.87	1.22	1.30	1.20	0.95	0.90	0.71*	0.88	0.46***	0.86	1.56	0.97	0.98
South-East	1.20	0.99	1.57**	1.69**	0.99	1.26	0.96	1.39	1.64*	1.25	1.04	1.20	1.53*	1.88**
South-West	1.19	1.20	1.19	1.52*	1.49*	1.00	1.09	1.14	1.31	1.05	0.95	1.10	1.57*	2.60***
East Anglia	1.17	1.57*	1.89*	2.27**	1.17	2.61*	1.23	2.17*	3.48**	1.65	1.34	1.52	2.22*	3.33**
The Midlands	1.11	0.91	1.25	1.53**	1.00	1.24	1.03	1.18	1.14	1.34	1.07	1.51*	1.63*	1.67*
The North	1.26**	0.95	1.47*	1.40*	1.02	1.77*	1.26	1.26	1.74**	1.31	0.98	1.09	1.38	1.79**
Wales	1.23*	0.81	1.81*	1.35	0.69	1.11	0.84	1.03	1.26	1.19	1.23	1.17	1.51	2.44**
Scotland	1.13	0.79	1.40*	1.72**	0.45***	0.96	0.62*	1.24	1.12	1.19	0.67	1.81*	1.13	1.26
N	8,961	8,126	7,683	7,354	7,016	6,720	6,418	6,248	6,031	5,864	5,679	5,528	5,412	5,287
Pseudo R^2	0.034	0.032	0.030	0.030	0.033	0.032	0.033	0.031	0.037	0.039	0.041	0.043	0.041	0.045

* The entries are odds ratios. In every wave response is modelled conditional on responding in all of the previous waves. The models in waves 2 to 5 used variables from wave 1; models in waves 6 to 10 used variables from wave 5; and models in waves 11 to 15 used variables from wave 10. The reference categories of the categorical independent variables in the table are male, non-white, good health, not a home owner, does not have a GCE or above degree, unemployed and others not present when interviewed, multi-person HH, household with no children, living in a house, no HH member is in employment in and London * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

Table 1.4 Response propensity models based on the AWA: modelling response in waves 5, 10 and 15.

	Response in wave 5 conditional on wave 1	Response in wave 10 conditional on wave 1 and 5	Response in wave 15 conditional on wave 1, 5 and 10
Female	1.24***	1.19**	1.30***
White	1.54*	1.61**	1.35
Age	1.08***	0.99	1.13***
Age-squared	0.99***	1.00	0.99***
Bad health	0.89	1.08	0.62***
Home owner	0.88	0.91	1.21*
Has GCE qualification or above	0.92	1.05	1.26**
Employed	1.37*	1.04	0.94
Others present when interviewed	1.11	1.09	1.19*
Single-person household	0.63*	1.18	0.89
Household with children	1.13***	1.07	1.15
Living in a flat	0.98	0.96	0.91
Living in other type of accommodation	0.68**	1.01	0.93
1 or 2 persons in employment in HH	0.83*	0.90	0.81*
3 or more persons in employment in HH	0.95	0.71*	0.89
South-East	1.34***	1.32*	1.37*
South-West	1.43***	1.33*	1.30*
East Anglia	1.99***	2.09***	2.03***
The Midlands	1.15	1.08	1.34*
The North	1.22*	1.40**	1.21
Wales	1.31*	1.02	1.49*
Scotland	1.16	0.79*	0.98
N	8,961	7,311	6,019
Pseudo R^2	0.038	0.029	0.040

* The entries are odds ratios. The model in wave 5 used variables from wave 1; the model in wave 10 used variables from wave 5 and the model in wave 15 used variables from wave 10. The reference categories of the categorical independent variables in the table are male, non-white, good health, not a home owner, does not have a GCE or above degree, unemployed and others not present when interviewed, multi-person HH, household with no children, there is at least one person aged 75+ in HH, living in a house, no HH member is in employment and London * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

We created two sets of weights to adjust for the longitudinal non-response at wave 15: standard weights (*SWs*) and alternative weights (*AWs*). The *SWs* were derived from the models in table 1.3, whereas the *AWs* were derived from the models in table 1.4. The derivation of the weights was based on equation 1.2 (for *SWs*) and 1.5 (for *AWs*) as discussed earlier. Also, as mentioned before, for our evaluation purposes, the *SWs* and *AWs* will be used in the analysis of savings and debt using the responding sample in waves 1, 5, 10 and 15. Thus, we are interested in *SWs* and *AWs* for a balanced panel from waves 1, 5, 10 and 15 (5,132 respondents). Table 1.5 and 1.6 show the distributions of these weights for both the *SWs* and *AWs*.

Table 1.5 presents the measures of central tendency and dispersion for both of the *SWs* and *AWs*. The distribution of the *AWs* is similar to the *SWs* in terms of the mean, median, and first and third quantiles weights, but the dispersion in the latter is greater. This is indicated by the larger standard deviation (1.11), bigger coefficient of variation (0.53) and the wider range of weights (0-17.59) in the set of *SWs*. This larger dispersion in the *SWs* is caused by the zero-value weights that the *SWs* reserve for a proportion of respondents who are present in the balanced panel in question but did not respond in all 15 waves. It is also possible that part of the larger variation in the set of *SWs* is due to the larger number of response propensity models in the SWA (14 model) by which more error components were modeled as opposed to the AWA (only 3 models).

Table 1.6 presents the frequency distribution for categorised weight values for both *SWs* and *AWs*. Excluding the first and last categories, for both sets of weights, respondents are rather evenly distributed across the weights values. In both sets of weights, most respondents have weights values between 1.51 and 2. The major difference between the values of the two sets of weights is that with the *SWs* 9.31% (or 478 cases) of our analysis sample will be assigned a weight of zero, whereas with the *AWs* all cases in the sample will have a non-zero weight. In addition, unlike the *AWs*, with the *SWs* more cases will have weights values above 2.

Based on these results, it can be concluded that, overall, the distributions of *SWs* and *AWs* are similar. However, the *SWs* have more variability than the *AWs*. Also, while the *AWs* enable us to use every sample member in our balanced panel sample in the analysis we intend to conduct, the set of *SWs* will reduce this sample by 9.31% by assigning a weight of zero to 478 respondents. Therefore, we can immediately expect that estimates resulting from the *AWs* to be more precise than the estimates resulting from the *SWs*.

Table 1.5 Distribution of the standard and alternative weights for a balanced panel from waves 1, 5, 10 & 15

	Mean	Std.dev	Min	Q1	Median	Q3	Max	CV	N
SWs	2.09	1.11	0	1.63	2.03	2.56	17.59	0.53	5,132
AWs	1.93	0.71	0.41	1.50	1.79	2.20	13.32	0.37	5,132

*Std.dev is the standard deviation. CV is the coefficient of variation, cv=std.dev/mean.

Table 1.6 Frequency distribution of the standard and alternative weights values

Categorised weights values	0	0.41-1.50	1.51-2.00	2.01-2.50	2.51-4.00	4.01-max
SWs	478 *(9.31%)*	500 *(9.74%)*	1,533 *(29.87%)*	1,239 *(24.14%)*	1,166 *(22.72%)*	216 *(4.21%)*
AWs	0 *(0%)*	1,341 *(26.13)*	1,974 *(38.46%)*	1,062 *(20.69%)*	667 *(13.00%)*	88 *(1.71%)*

*The entries are the number of respondents that falls in the given category of weights values. Respondents are from a balanced panel from wave 1, 5, 10 and 15.

1.6.2 Modelling savings and debts

The BHPS provides detailed information on savings and debts at the individual level in our waves-combination of interest. In each of the waves, respondents were asked if they have money in savings and whether they owe money. If respondents have money in savings and/or owe money, they are then asked to state these amounts. This setting permits two main sets of outcome variables which were used in the analysis: (a) Dichotomous: these are two variables, one indicates whether an individual has savings or not and the other indicates if the individual is in debt (1=having savings, 0=having no savings; and 1=having debt, 0=having no debt); (b) Continuous: these are two variables reflecting the amounts of savings and debts.

The main independent variables used in the analysis are year of data collection (wave), gender, race, age, annual income, marital status, labour force status, whether respondents have their own children, financial status and household size, as these variables are important in predicting both the existence and level of wealth (Kan and Laurie, 2010). Gender, race and age were also used in the response propensity models; however, these three variables are fundamental in most analyses of social science processes.

We analysed a balanced panel of those aged 16+ from waves 1, 5, 10 and 15 (5,132 respondents) with our two sets of weights. Recall that the standard weights will reduce the size of the panel to 4,654. The analysis was carried out in STATA. The data was introduced as a panel data set so that the multiple observations per person are linked to one case rather than being treated as different cases. Since the proportion of missing values for the amounts of savings and debt were high (over 19% for saving, and over 7% for debt), the missing values were imputed to reduce any bias that might be brought into the analysis. The imputation was carried out using the common *Hot-deck* (built-in) command in STATA. The *Hot-deck* procedure performs random imputation, which involves categorizing the respondents in the sample into similar subgroups based on a number of variables. The variables used here were gender, age group, race and household size. These variables were chosen for the imputation because they are some of the best predictors of the amounts of savings and debt (Kan and Laurie, 2010). Missing data for respondents in any subgroup are randomly replaced with comparable data from respondents in the same subgroup. The values were only imputed for those who reported having savings or are in debt. Those who reported that they do not have money in savings or are not in debt, their corresponding amounts of savings and debt were set to 0. The effect of this imputation is included in appendix A.2.

Two types of analyses were done:

First, descriptive statistics: with multiple observations per person type of data (panel data) interesting statistics in the context of analysing debt data are the transition probabilities into debt. Thus, for debt (1= in debt and 0= not in debt), we estimated a transition matrix using the standard and alternative weights. Additionally, we estimated weighted proportions of those who have savings and debt using the standard and alternative weights separately. The results from all of these analyses are compared and discussed in the next section.

Second, multivariate analysis: the structure of the data (each case is linked to multiple observations collected at different times) allows for the application of panel data models such as random effects or fixed effects. Two panel data models, namely, random effects logistic regression models were used to estimate the determinants of having money in savings or being in debt respectively. However, each model was estimated twice using the two different sets of weights. Similarly, to model the amounts of savings and debts, two random effects OLS regression models were estimated in which every model was estimated two times using the two sets of weights. We used random effects models, rather than fixed effects, because some of our explanatory variables are time-invariant variables.

With two different sets of longitudinal weights, eight models were estimated as each set of weights was used to estimate all of the four models separately.

However, clustering and stratification were not specified in the analysis, as STATA – like many statistics software – does not allow this while estimating panel data models.

Therefore, this may lead to under-estimating the standard errors of estimates in this analysis. Nonetheless, although this may not result in precise estimates, any differences between the produced estimates will be due to the difference between the two sets of weights since the modeling strategy is held constant. Thus, it is possible to draw a conclusion on whether the two sets of weights have different impacts on the substantive analysis.

1.7 Results

1.7.1 Descriptive results

Table 1.7 presents two 2x2 transition probability matrices for debt. The first matrix was estimated using *SWs*, whereas the second matrix was estimated using *AWs*. Both matrices were estimated as follows:

We have two observed debt transitions (between wave 5 and 10, and between wave 10 and 15). The overall transition probability matrix for debt (between wave 5 and 15), is therefore the sum of the two component transition matrices. We calculated the number of weighted transitions between wave 5 and 10, and between wave 10 and 15 amongst all categories of debt. We then combined (summed) the equivalent transition numbers before calculating the probabilities based on the summed numbers.

The first matrix was estimated using the standard weighs (*SWs*), whereas the second matrix was estimated using the alternative weights (*AWs*). Overall, both of the *SWs* and *AWs* result in similar transition probabilities. For example, for the transition matrix estimated with the *SWs*, those who are not in debt transition into debt with a probability of

0.179, whereas the *AWs* indicate that this probability is 0.193. Also, the *SWs* show that those who are in debts may clear this with a probability of 0.427; meanwhile the *AWs* show that the equivalent probability for this is 0.442. While all equivalent probabilities are similar in general, each pair of equivalent probabilities indicates a difference of, mostly, 1% between the two probabilities. Such differences may, sometimes, have a significant impact on the interpretation of the results and, consequently, on the pertinent decision-making process. Thus, these differences indicate that the *SWs* and *AWs* are slightly different from each other. The additional 478 cases that are associated with the *AWs* change the transition counts amongst debt statuses (to higher numbers of transitions) and therefore resulted in, slightly, different transition probabilities compared to the *SWs*.

Table 1.7 Debt transition probability matrix.

	Using *SWs*		Using *AWs*	
	Not in debt	In debt	Not in debt	In debt
Not in debt	.821 *(.0047)*	.179 *(.0032)*	.807 *(.0045)*	.193 *(.0031)*
In debt	.427 *(.0036)*	.573 *(.0028)*	.442 *(.0035)*	.558 *(.0026)*

* The rows reflect the initial statuses of debt, and the columns reflect the final statuses. The numbers in brackets are the standard errors.

Table 1.8 presents weighted proportions of those who reported having savings and debt respectively and the corresponding standard errors of these proportions. The proportions were calculated using *SWs* and *AWs* separately. For savings, both of the *SWs* and *AWs* show that 43% of respondents have these. However, the standard error of the proportion is larger with *SWs* (.0038) than with *AWs* (.0022). As for debt, the relevant proportions are 35% and 36% with *SWs* and *AWs* respectively indicating a small difference (of 1%)

between the two proportions. The standard error of the proportion of those who have debt is also larger with *SWs* (.0037) than with *AWs* (.0021).

These results confirm our expectations regarding differences in terms of precision between the resultant estimates. While both sets of weights result in similar proportions of those who have savings and debt, the corresponding standard errors of these proportions are smaller with the *AWs* than with *SWs*. This outcome is sensible as the sample size associated with the *AWs* is larger, and also because the *AWs* have smaller variance compared to the *SWs*.

Table 1.8 Weighted proportions of those who have savings and debt.

	Savings		Debt	
	%	SE	%	SE
With standard weights	43	.0038	35	.0037
With alternative weights	43	.0022	36	.0021

* SE is the standard error of the given proportion.

1.7.2 Results from the multivariate analyses

Possession of Savings and Debts

Table 1.9 presents the results of the random effects logistic regressions of modelling the possession of savings and debt respectively. The table presents odds ratios and their standard errors. The models were estimated using a balanced panel from waves 1, 5, 10 and 15. Each model was estimated with *SWs* and *AWs*. As can be seen from all models, the possessions of savings and debts are highly associated with financial situation, income and labour force status. For example, higher income is positively associated with having both savings and debts ($\hat{b}_{Save,SWs}$ = 1.027, p< 0.001; $\hat{b}_{Save,AWs}$= 1.028, p< 0.001; $\hat{b}_{Debt,SWs}$ = 1.011, p< 0.001; $\hat{b}_{Debt,AWs}$= 1.011, p<0.001), meanwhile, those who are out of the labour force are less likely to have savings and debts than those who are employed ($\hat{b}_{Save,SWs}$= 0.618, p<0.001; $\hat{b}_{Save,AWs}$= 0.603, p<0.001; $\hat{b}_{Debt,SWs}$ = 0.229, p<0.001; $\hat{b}_{Debt,SWs}$= 0.245, p<0.001).

Looking at the magnitudes of the coefficients resulting from the *SWs* and *AWs*, for both of our two substantive outcomes, one can immediately notice that the two models are approximately the same. However, there are a few differences:

Unlike the *AWs*, using the *SWs* reduced the sample size, as expected, to 4,654 for both savings and debt models as 478 respondents are assigned a weight of zero. As a result, the standard errors of many coefficients in the *AWs* models are smaller. Consequently, in these models, the significance levels of some of the coefficients are increased in comparison with their equivalent coefficients in the models estimated with *SWs*.

For instance, in the models of possession of savings, the standard errors of 'living with a partner' and 'member of a large household' dropped from (*.043*) and (*.037*) to (*.035*) and (*.030*) respectively as a result of changing the weights from *SWs* to *AWs*. Consequently, the significance levels of these variables are increased ($\hat{b}_{SWs}$ = 1.084, p< 0.05, $\hat{b}_{AWs}$= 1.095, p<0.01; $\hat{b}_{SWs}$ = 0.882, p< 0.01, $\hat{b}_{AWs}$= 0.864, p< 0.001) respectively.

As for debts, the differences appear with the 'year 2005' and 'member of a large household'. While the significance of 'member of a large household' is increased with

AWs ($\hat{b}_{SWs}$ = 1.093, p< 0.05, $\hat{b}_{AWs}$= 1.134, p< 0.01), 'year 2005' only appears significant with AWs ($\hat{b}_{SWs}$ = 1.065, p> 0.05, $\hat{b}_{AWs}$= 1.099, p< 0.05).

These results clearly show the effect of the loss of 478 respondents from the analysis, which is associated with the SWA. These 478 respondents were picked up by the AWA and therefore changed the sample used in the estimation, and consequently produced more precise estimates. Also, another contributing factor to producing more precise estimates with the AWA is the fact that there is less variation in the AWs than in the SWs.

Table 1.9 Random effects logistic regression models of possession of savings and debts.

	Having Savings		Having Debts	
	Using standard weights (*SWs*)	Using alternative weights (*AWs*)	Using standard weights (*SWs*)	Using alternative weights (*AWs*)
Year 2000	0.992 *(.033)*	1.002 *(.033)*	1.001 *(.036)*	1.032 *(.036)*
Year 2005	1.017 *(.034)*	1.000 *(.034)*	1.065 *(.038)*	1.099 *(.035)**a
Female	1.106 *(.043)***	1.032 *(.042)***	1.047 *(.045)*	1.048 *(.045)*
White	1.047 *(.226)*	1.007 *(.106)*	2.014 *(.226)****	2.021 *(.226)****
Aged 26 to 45	0.973 *(.077)*	0.981 *(.077)*	0.735 *(.058)****	0.757 *(.057)****
Aged 46+	1.176 *(.063)***	1.185 *(.063)***	0.318 *(.026)****	0.339 *(.026)****
Living with a partner	1.084 *(.043)**	1.095 *(.035)***a	1.042 *(.046)*	1.068 *(.046)*
Financially okay	0.668 *(.025)****	0.680 *(.025)****	1.531 *(.065)****	1.530 *(.065)****
Financially struggling	0.195 *(.008)****	0.200 *(.008)****	2.152 *(.098)****	2.163 *(.097)****
Member of a large HH	0.882 *(.037)***	0.864 *(.030)****a	1.093 *(.047)**	1.134 *(.041)***a
Has dependent children	0.790 *(.037)****	0.786 *(.037)****	1.153 *(.055)***	1.146 *(.055)***
Annual income/1000	1.027 *(.002)****	1.028 *(.002)****	1.011 *(.001)****	1.011 *(.001)****
Unemployed	0.543 *(.030)****	0.567 *(.030)****	0.620 *(.035)****	0.622 *(.035)****
Out of the labour force	0.618 *(.028)****	0.603 *(.028)****	0.229 *(.012)****	0.245 *(.010)****
Has a second job	1.374 *(.080)****	1.331 *(.080)****	1.518 *(.092)****	1.473 *(.092)****
N	4,654	5,132	4,654	5,132
σ	1.24	1.22	1.36	1.34
ρ	0.31	0.31	0.35	0.35

Note: The entries are odds ratios. The numbers in brackets are the standard errors. [a] indicates a difference in the significance level between the equivalent coefficients. The reference categories of the dependent variables are having no savings and having no debts respectively. The reference categories of the categorical independent variables in the models are year 1995, male, non-white, aged 16 to 25, not living with a partner, having a good financial situation, having a small household, has no dependent children, employed and has no second job respectively. σ is the standard error of random effects (sigma u). ρ is the percentage of the total variance that is due to differences between units, ρ=sigma^2u/(sigma^2u +sigma^2e). *p< 0.05, ** p< 0.01 and *** p<0.001.

Amounts of Savings and Debts

The models in table 1.10 show the results of OLS random effects models of the amounts of savings and debts using the responding sample at wave 1, 5, 10 and 15. Both the amounts of savings and debt were modelled using *SWs* and *AWs* separately.

Similar to the possessions of savings and debts, the results here show that the levels of savings and debt are also highly associated with financial situation, income and labour force status. For example, higher income is positively associated with the amounts of both savings and debts ($\hat{b}_{Save,SWs}$ = .0028, $p < 0.001$; $\hat{b}_{Save,AWs}$= .0027, $p < 0.001$; $\hat{b}_{Debt,SWs}$ = .045, $p < 0.001$; $\hat{b}_{Debt,AWs}$= .042, $p < 0.001$). Also, those who are out of the labour force have lower amounts of savings and debts than those who are employed ($\hat{b}_{Save,SWs}$= -0.012, $p < 0.001$; $\hat{b}_{Save,AWs}$= -0.0216, $p < 0.001$; $\hat{b}_{Debt,SWs}$ = -0.858, $p < 0.001$; $\hat{b}_{Debt,SWs}$= -0.799, $p < 0.001$).

Turning to the effect of the different weighting approaches on the models, for both savings and debt, the *SWs* and *AWs* gave similar results. However, differences in terms of precision were found. For both saving and debt, the model estimated with *AWs* results in smaller standard errors for most coefficients in comparison with the model estimated with *SWs* as can be seen in table 1.10. This, in turn, increased the significance of a few variables in the models estimated with *AWs*. For the amount of saving model, for instance, ‘female’ appears significant only if the model is estimated with *AWs* ($\hat{b}_{SWs}$ = -.0061, $p > 0.05$, $\hat{b}_{AWs}$= -.0060, $p < 0.05$). Also, ‘living with a partner’ and ‘has dependent children’ are more significant with *AWs* than with *SWs* ($\hat{b}_{SWs}$ = .0082, $p < 0.05$, $\hat{b}_{AWs}$= .0106, $p < 0.01$; $\hat{b}_{SWs}$ = -.0153, $p < 0.01$, $\hat{b}_{AWs}$= -.0166, $p < 0.001$) respectively.

Focussing on the amount of debts models, ‘has dependent children’ does not appear significant with *SWs*; whereas, using the *AWs* shows that this variable is significant at the level of *0.05* ($\hat{b}_{SWs}$ = 0.257, $p > 0.05$, $\hat{b}_{AWs}$= 0.212, $p < 0.05$).

These results are consistent with the results of modelling the possessions of savings and debt indicating that the results associated with the *AWs* are more precise, and hence implying that the AWA is more efficient in dealing with non-response than the SWA when non-monotonic response pattern applies.

In sum, based on our descriptive and multivariate analyses, the SWA and AWA result in similar estimates. However, as a consequence of implementing different methodologies to identify response in the two approaches, weighted estimates may be affected differently. On the SWA part, some respondents are unnecessarily lost. As a result, the sample size used in the weighted analyses associated with the SWA is smaller leading to larger standard errors of estimates resulting from these analyses. Consequently, the importance of some factors in analyses that use weights from the SWA may be under estimated.

On the other hand, the AWA identifies response by aiming to avoid losses of respondents whose data can be used in the given analysis. Thus, the sample size used in weighted analyses associated with the AWA is larger. This, in turn, produces estimates with smaller standard errors in comparison with the SWA. Consequently, some estimates appear more significant if estimated with AWs instead of SWs.

Table 1.10 Random effects OLS regression models of the amounts of savings and debts.

	Savings		Debts	
	Using standard weights (*SWs*)	Using alternative weights (*AWs*)	Using standard weights (*SWs*)	Using alternative weights (*AWs*)
Year 2000	.0054 *(.004)*	.0043 *(.003)*	0.199 *(.140)*	0.165 *(.127)*
Year 2005	.0006 *(.004)*	.0002 *(.003)*	-0.132 *(.140)*	-0.118 *(.127)*
Female	-.0061 *(.004)*	-.0060 *(.003)**a	-0.521 *(.128)****	-0.554 *(.118)****
White	-.0090 *(.010)*	-.0095 *(.010)*	-1.256 *(.350)****	-1.009 *(.319)****
Aged 26 to 45	.0007 *(.008)*	.0018 *(.007)*	-0.518 *(.309)*	-0.326 *(.280)*
Aged 46+	.0001 *(.009)*	.0043 *(.008)*	-1.265 *(.311)****	-1.104 *(.282)****
Living with a partner	.0082 *(.004)**	.0106 *(.003)***a	0.006 *(.141)*	0.004 *(.129)*
Financially okay	-.0221 *(.004)****	-.0216 *(.004)****	0.096 *(.145)*	0.101 *(.133)*
Financially struggling	-.0419 *(.004)****	-.0412 *(.004)****	0.706 *(.152)****	0.614 *(.140)****
Member of a large HH	-.0021 *(.004)*	-.0040 *(.004)*	-0.249 *(.159)*	-0.166 *(.146)*
Has dependent children	-.0153 *(.005)***	-.0166 *(.004)****a	0.257 *(.179)*	0.212 *(.164)**a
Annual income/1000	.0028 *(.001)****	.0027 *(.001)****	0.045 *(.005)****	0.042 *(.004)****
Unemployed	-.0119 *(.005)**	-.0124 *(.005)**	-0.686 *(.197)****	-0.612 *(.178)****
Out of the labour force	-.0212 *(.005)****	-.0216 *(.004)****	-0.858 *(.165)****	-0.799 *(.152)****
N	4,654	5,132	4,654	5,132
σ	0.07	0.07	1.14	1.33
ρ	0.15	0.15	0.03	0.04

Note: The numbers in brackets are the standard errors. [a] indicates a difference in the significance level between the equivalent coefficients. The reference categories of the dependent variables are having no savings and having no debts respectively. The reference categories of the categorical independent variables in the models are year 1995, male, non-white, aged 16 to 25, not living with a partner, having a good financial situation, having a small household, has no dependent children, employed and has no second job respectively. σ is the standard deviation of the random effects (sigma u). ρ is the percentage of the total variance that is due to differences between units. * $p< 0.05$, ** $p< 0.01$ and *** $p<0.001$.

1.8 Conclusion

Allowing for non-monotonic response pattern by attempting data collection from sample members every wave regardless of their previous response statuses is beneficial in a number of dimensions. Apart from collecting data that can be useful to adjust for missingness, it assists identifying whether sample members who did not provide response for one or more waves are still eligible for the survey. However, it may result in different responding samples in different wave-combinations. Some of these combinations of waves may be used separately for analysis.

In this chapter, we evaluated the SWA in the analysis of a subset of waves-combination when non-monotonic response pattern applies. We also introduced and evaluated an

AWA which is, unlike the SWA, creates a non-zero weight for every sample member in the responding sample in the wave-combination in question. Additionally, we distinguished between the two weighting approaches with respect to the number of sets of weights. The strategy of the SWA is to create a single set of longitudinal weights. The AWA, on the other hand, allows creating a number of subsets of weights for a number of wave-combinations. Each of these wave-combinations may be considered because it contains data on a particular social phenomenon, and hence may be used for analysis separately.

By analysing wealth data from waves 1, 5, 10 and 15 of the BHPS, we found that using *SWs* results in a loss of 9.31% (478 respondents) of our analysis sample as their corresponding weight values in the *SWs* are zeros. In contrast, using the *AWs* assigned a non-zero weight for all the responding sample members in waves 1, 5, 10 and 15, and therefore, the full responding sample was used in the analysis.

It is obvious that the SWA is disadvantageous when the analysis is restricted to a subset of wave-combinations as it allocates zero weights for some respondents in this case. The problem is due to the methodology of the SWA in identifying response which does not take into account the fact that the response pattern is non-monotonic. As a result, the SWA only identifies one group of sample members as responding (those who responded in all waves). In return, the AWA recognises that, with a non-monotonic pattern of response, the responding sample may differ across different wave-combinations. Thus, it creates its weights by identifying response as responding in the wave-combination used in the analysis.

Since wave-combinations that contain specific measures are likely to be used for analysis separately, and that the SWA may not be the best option in this case, we recommend that the AWA should be used to create subsets of weights for analyses restricted to these combinations of waves.

Our findings, however, suggest that, even when the analysis is restricted to a subset of wave-combinations, the SWA and the AWA result in similar estimates. The difference is that the SWA may result in less precise estimates as a consequence of excluding a proportion of respondents from the analysis sample which can be avoided by using the AWA.

It is important to notice that the findings here do not suggest that the AWA is a complete replacement of the SWA. As most longitudinal analyses are based on a balanced panel from all of the conducted waves, the SWA remains useful in many types of analyses. But it could be well supported if a limited number of additional subsets of weights are created based on the AWA introduced in this chapter. Our findings here support this argument.

The extra sets of weights will serve as strong alternatives when analysis is restricted to wave-combinations that collect data on the same subject. Thus, for survey organisations that include a specific module of questions in particular combinations of waves, their weighting can be largely improved from creating extra subsets of weights for these particular wave-combinations in addition to the set of *SWs*.

On the data analysts' part, those who would like to construct estimates using a specific set of measures that are included in a particular waves-combination, but also like to adjust for non-response in their estimates, the *SWs*, as mentioned, may be suboptimal. However, many analysts will just assume that the weights provided by the survey organisation will reduce non-response bias without affecting other aspects of their analysis, such as reducing the size of their analysis sample. As demonstrated in this chapter, with the SWA, this is not always the case.

Even data users who can realise that the SWA is suboptimal in some analyses, but they are concerned about non-response bias, will be forced to use the *SWs* if these are the only weights offered in the data file. Even though this may be at the cost of reducing their analysis sample and hence reducing precision in their estimates. Consequently, those who cannot afford to increase the variance of their estimates, but at the same time they do not want to risk their estimates with potential non-response bias would face a rather tricky decision with the single option of *SWs*.

Therefore, with the additional subsets of weights, survey organisations will provide a rare opportunity for weighted analyses that are only based on topic-specific wave-combinations. When using the *AWs* instead of the *SWs*, analysts can ensure that the precision of their estimates is improved, while the estimates remain relatively unchanged as suggested by the evidence in this chapter.

It may of course be the case that survey organisations realize that the SWA has its limitation with regard to non-monotonic response pattern. Still, the SWA is preferred. From a critical viewpoint, the issue is that many individuals, even within some academic survey organisations, are not convinced that a considerable proportion of data users will use any created weights. Moreover, most substantive analyses use a balanced panel from all waves for which the *SWs* are suitable. Thus, it is not necessary spend extra time creating additional subsets of weights that might not be distinguished, by most data users, from the *SWs* let alone the fact that they may not be used.

However, such attitude is - to an extent - hypothetical, and is in clear contradiction with the principles of providing comprehensive and reliable data sets that academic survey organisations adopt. Survey organisations carry the burden of creating and releasing non-response weights. A reasonable assumption that has to be made, then, is that some analysts are concerned about non-response bias, and if they decide to use adjustments offered by the data providers, they may want to use the best available alternatives.

Topic-specific wave-combinations are typically few, and are already defined by the survey design. Additionally, it is not difficult to modify the weighting model in the SWA to create the extra sets of weights. In return, it is advantageous and in line with the development in weighting schemes that leading longitudinal surveys implement.

By using waves 1, 5, 10 and 15 from the BHPS, it was shown that the AWA could prevent the loss of 9.31% of the analysis sample. In similar combinations of waves in other surveys, the effect of this method could be more significant. This is particularly more likely in surveys with larger number of waves. Thus, for surveys that are designing a long-life panel, such as Understanding Society in the UK, this approach of weighting is worth considering.

Aside from designing subsets of weights for wave-combinations that collect data on the same subject, survey organisations may use other criteria to decide on potential combinations of waves that require *AWs*. For example, to enhance the accuracy of survey estimates, survey organisations sometimes add extra information to the original sample.

For instance, two samples (from Scotland & Wales) were added to the BHPS in wave 9. Also, an additional sample (from Northern Ireland) added at wave 11. Thus, for the BHPS, providing subsets of weights for waves 9 onwards, and 11 onwards might be of interest. Other criteria were put forward by Lynn and Kaminska (2010).

Although each suggested criterion may be useful, at present there is no approach that guides all of – or at least some of - these criteria simultaneously for a rational selection of wave-combinations for weighting. Instead, each criterion, if used, will be used solely. Meanwhile, weighting can benefit greatly by combining all or some of these considerations. Thus, a challenge for future research is to unite all these considerations into one constructive method that would guide the process of selecting wave-combinations.

In any case, the choice of specific waves-combination for weighting should be guided by a rule that takes into account two issues:

(a) The subsample drawn for analysis from any chosen combination of waves should be considerably different (in size and composition) from the samples used in the SWA.

(b) The selected combination of waves should be usable for analysis that achieves the objectives of the survey.

CHAPTER 2

Unknown Eligibility whilst Weighting for Non-response: the Puzzle of who has Died and who is still Alive?

2.1 Introduction

Household longitudinal surveys follow respondents over time and continue gathering data from them, which not only provide an opportunity for cross-sectional analysis on the data collected at each wave, but also allows the analyses of gross change. These features of longitudinal studies lead to a thorough understanding of dynamic populations, which cannot be achieved by conducting cross-sectional surveys. However, for this to be achieved, a well-designed representative sample is required.

The sample selected for a panel study is usually designed to represent the population of interest at the start of the survey. Surely, the population of interest changes over time as people are born, immigrate, die and emigrate (Lynn, 2011). Through these changes, some units will leave the study population, while other units will enter this population. For example, if the study population is defined as ‘those who are alive and resident in households in the state where the survey is conducted’ (a common definition of eligibility for household panel surveys), those who die will no longer be part of the population (ineligible) and those who are newly born will enter the defined population.

Thus, the sample should also be modified over time to maintain representativeness of the population of interest during the course of the study. New eligible members may join the sample through a specific mechanism that could be established by the sample design. For example, to represent the new births in the population, the survey might establish a rule that new born children of any eligible female sample member will be added to the sample as eligible sample members[7]. Accordingly, survey researchers can, relatively, control the

[7] This is how new births are represented in the sample of Understanding Society in the UK (Lynn, 2011).

system by which new members should join the sample, and hence new eligible sample members can be known.

In turn, some sample members may die or move out of the scope of the survey, in which case they become ineligible for survey administration. These cases should be identified as they are no longer part of the population of interest, otherwise they might cause a number of complications in the survey as shall be discussed in this chapter. However, for some subgroups in the sample, identifying those sample members who become ineligible can be a challenging task. In most samples of longitudinal surveys, there is a proportion of sample members whose eligibility is unknown. A major reason for this is that survey organisations can lose track of sample members if, for instance, they change their residence address and contact details without informing the survey office. These cases, and any others where contact cannot be established, result in non-response. Non-response through non-contact obstructs the identification of the eligibility status of sample members, since little information about them is available.

This is particularly a dilemma at waves subsequent to the last successful contact attempt as information about those who have not been successfully contacted for some waves may not be available at all. For example, consider a 10-wave survey, where sample members are eligible if they are alive. If, after wave 1, the survey could not re-establish contact with a group of sample members, some of whom may have been aged 90 or more at the start of the survey, it would be rather tricky to classify them as either eligible or ineligible by the time of wave 10. A plausible question in these circumstances is whether some of the sample members in question are in fact ineligible (deceased). Therefore, determining the eligibility status for sample members whose eligibility is unknown is an important challenge for survey researchers.

Unknown eligibility raises a number of practical concerns in longitudinal surveys. The major concerns are:

(a) Disturbance of the calculation of survey quality measures: accurate calculation of important measures of the survey quality such as the response rate, contact rate and co-operation rate can be affected if eligibility is unknown for a considerable proportion of sample members. In order to calculate the quality rates, it is important to identify the number of eligible sample members in the sample as it determines the base on which these rates are calculated. Thus, appropriate classification of those whose eligibility is unknown as either eligible or ineligible may be necessary. In this chapter, we briefly discuss the effect of unknown eligibility on the calculation of the response rate, contact rate and co-operation rate, but our main focus will be on

(b) Potential distortion of non-response weighting: unknown eligibility may negatively affect non-response weighting. This is particularly likely if the treatment of cases whose eligibility is unknown is done as in the standard weighting approach (SWA). In the SWA, sample members whose eligibility is unknown are assumed to be eligible, and are therefore included in the process of estimating the weights for the responding cases. Consequently, if some of the cases whose eligibility is unknown are actually ineligible, the weights will encompass influence of units that are not part of the population of interest in the weighted estimates. In turn, weighted estimates may be biased.

In this chapter we investigate an alternative weighting approach (AWA). The AWA will adjust the standard weights (*SWs*), resulting from the SWA, to reduce the effect of potentially including ineligible cases during weights creation. The weights resulting from the adjustments made by the AWA will be used as the alternative weights (*AWs*). To make its adjustments, the AWA will use information on eligibility from the study population, which could be available from an external source. Details on the logic and methodology of the AWA are given in section (2.7). We use the BHPS sample to carry out the investigation. For the assessment of the AWA, we analyse data on subjective health status using the *SWs* and the *AWs* and compare the results. To check the sensitivity of the results generated from the AWA, we introduce a second alternative approach. The latter involves an imputation procedure to deal with unknown eligibility during weighting. Full explanation of the second alternative approach is also given in section (2.7). However, our attention is given to the AWA, and the second alternative is mainly introduced to provide a sensitivity analysis.

Aside from survey-specific characteristics of ineligible sample members, the most common characteristics of being ineligible for a survey are: death, moving out of the geographical area that is covered by the survey (emigration) and being institutionalised – such as going to prison or residing in a military base -. In all these cases, establishing contact may not be possible[8]. However, in this chapter, we limit our investigation to death. This is because of three reasons. First, unlike other forms of ineligibility, in most countries, information on mortality is usually well documented and available for public

[8]Contact is sometimes possible after death. This may happen, for example, if contact is established with another household member who can inform the interviewer of the death of the sample member in question. However, contact may not be established in circumstances where the deceased sample member had previously lived alone.

use. Often such information is recorded (by age and gender) for the whole population and can be obtained from a single external source such as national statistics institutes. In contrast, data on those who emigrate or move from a household to an institution, are not necessarily reported, and hence may not be accurately recorded by a single source. Furthermore, such data, if recorded, are not typically available for access by members of the public, and even if they are released they are often in aggregate levels which may not be useful for a thorough investigation.

Second, death is a larger source of ineligibility compared to emigration and institutionalisation (Watson, 2014).

Third, emigration and institutionalisation are more complex forms of ineligibility than death. With emigration and institutionalisation, it is possible for eligibility status of individuals to change (possibly multiple times) between ‘eligible’ and ‘ineligible’ at different time points causing difficulties in terms of estimating valid eligibility statistics. On the other hand, the terminal nature of death makes it a simple form of ineligibility. If a person is ineligible through death, they remain ineligible, which makes the use of mortality information relatively reliable.

Thus, our AWA will focus on dealing with the group of dead sample members amongst those whose eligibility is unknown. The terms ‘dead’ and ‘ineligible’ (or alive and eligible) may be used interchangeably here.

Additionally, the chapter discusses the concepts of eligibility and unknown eligibility in survey research in general; and reviews a number of methods used to estimate eligibility rates amongst those whose eligibility is not known. Though, our major focus will be on dealing with unknown eligibility in the weighting context.

2.2 Eligibility

The objective of sample surveys is to make inference about a population based on information obtained from a sample. Usually, the population of interest is defined precisely according to specific characteristics. Sample units whose characteristics match the characteristics of the population of interest are referred to as eligible sample units. Defining eligibility is a crucial step in every survey. Conditions for being eligible vary between surveys, depending on the aim and objectives of the survey. In some surveys, the definition of eligibility is linked to a certain period or point in time. For example, in a survey of smokers, if being eligible is defined as being a smoker, the survey organisation should link the definition of eligibility to a specific time period, as individuals may start or stop smoking during the data collection period.

Since the only usable data for analysis are collected from eligible sample units, ineligible cases are dropped from the sample, and as a result, the sample size is then reduced. Thus, especially in cross-sectional surveys, it is advantageous to increase the eligibility rate, by, for example, pre-screening the sample units before selecting the sample. This is because, at the sampling stage, sometimes it is difficult for the survey researcher to spot some of the undesirable or ineligible cases (cases that are not part of the population of interest) in the sample frame. For instance, in random dialling digit surveys the sample frame may contain non-working numbers (i.e. ineligible); however, it might be impossible to know this unless a contact attempt is made (Groves *et al*, 2004).

In contrast, in longitudinal surveys, pre-screening the sample may not be of great benefit in the long term. This is because individuals' characteristics that match the characteristics of the population of interest can change overtime, introducing the possibility that sample members may be part of the population in earlier waves but not in later waves. For example, if being eligible in a survey is defined by living in the country where the survey is conducted, some sample members may leave the country after participating in a number of waves, and, as a result of this, they become out of the scope of the population of interest. This complexity demonstrates that dealing with eligibility in longitudinal surveys is more problematic.

Although it is cost effective if ineligible sample members are identified before they are issued to an interviewer, often ineligible units are not identified as such until the interviewer makes contact and finds out that a sample member is ineligible.

2.3 Unknown eligibility

The term 'unknown eligibility' is used to refer to the status where there is not sufficient information about a sample member to allow them to be identified as either an eligible or ineligible after the data collection stage is completed.

The most common outcomes of any contact attempt are: sample member is ineligible, completed interview, refusal or non-contact. In the case of ineligible sample members and completed interviews eligibility is defined. However, with non-response, which occurs through refusal or non-contact, information about non-respondents is very limited and sometimes not available. Therefore, the survey researcher may be able to identify eligibility for some of the non-respondents, but for a substantial proportion, eligibility will remain unknown. With refusals however, it is possible sometimes, particularly in household panel surveys, to identify eligibility. This is because eligibility in most panel studies is mainly defined as being alive and resident. Thus, although the survey fails to gain cooperation from those who refused to participate, it is possible to classify them as eligible sample members.

In panel studies, unknown eligibility can be resolved in the case of wave non-response, where sample members are not present for at least one wave, but they resume participation at some point during the course of the survey. In this case, information related to eligibility status during the period of absence can be collected in the current interview. In turn, a special case of unknown eligibility occurs through attrition. Attrition is the permanent dropout from a longitudinal survey after having participated at previous points of data collection (Chang, 2010). In this case, the survey researcher is unable to identify the eligibility status of sample persons even though they were eligible when they gave their last interview. Despite the use of different strategies to minimize attrition (McGonagle *et al*, 2011; Laurie and Lynn, 2009; Laurie *et al*, 1999), in some cases it is impossible to retain survey participation.

In panel surveys, some of the non-contacted sample members will in fact have died and if these deaths are not reported to the survey organisation, deceased sample members will be classified as sample members whose eligibility is unknown. In consequence, this may turn the process of maintaining the representativeness of the sample over time into a challenging task since decisions about those whose eligibility is unknown cannot be made easily.

Therefore, and particularly in household panel surveys, it is very useful if eligibility status of sample members whose eligibility is unknown, but are actually deceased (ineligible) is estimated as accurately as possible. This will benefit both the calculation of response rates and non-response weights as will be explained in sections 2.5 and 2.6 respectively. While it is challenging to estimate eligibility at the case level for cases whose eligibility is unknown, a number of methods can be used to estimate proportions of eligible cases. These are discussed next.

2.4 Methods for estimating the eligibility rate amongst cases of unknown eligibility

There are several methods that are usually used to estimate the rate of cases of unknown eligibility that are actually eligible ‘e’. Most of the literature in this area assumes a random digit dialling survey (RDD) (Smith, 2003) as this is a context in which typically eligibility cannot be established for a high proportion of sample cases. Therefore, some of the methods are RDD specific.

Minimum and maximum allocation: this is a simple method in which ‘e’ is assumed to be either 0% or 100% of the cases of unknown eligibility (Lessler and Kalsbeck, 1992; Smith 2003). Accordingly, more than one response rate can be produced. Taking ‘e’ as 0%, gives the maximum possible response rate while substituting ‘e’ as 100% produces the lowest response rate. This will be shown in the next section when the calculation of the response rate is discussed. However, one can obtain a range of response rates by varying the values of ‘e’ from 0% to 100% before choosing a plausible value that does not inflate the rate.

Proportional allocation[9]: this method assumes that 'e', among the cases of unknown eligibility, is the same as among the cases whose eligibility is known (Frankel, 1983; Lessler and Kalsbeck, 1992; Smith, 2003; Barron, Khare and Zhao, 2008). Smith (2003) indicates that proportional allocation is conservative as it produces a high value of 'e', but it might produce a biased estimate of 'e' because the assumption that the eligibility rate among the unobserved sample is the same as among the observed sample is unlikely to hold true.

Survival analysis: this method is the standard survival analysis method in which the number of contact attempts is used to estimate the eligible cases among the cases of unknown eligibility (Frankel *et al*, 2003). This method is considered to be a better approach to estimating 'e', since it uses more information from the sample than the other methods. However, it cannot be asserted that the statistical assumptions of survival analysis are properly met (Smith, 2003).

RDD specific methods: there are a few methods used in random digit dialling surveys to estimate the eligibility rate among the unknown eligibility cases. The most commonly used of these are: allocation based on disposition codes and contacting telephone business offices. Under the disposition codes allocation approach, the outcome of the call attempt is used to identify whether a case is eligible or not (Smith, 2003). For example, in a survey, a researcher might establish a rule that all of the phone numbers with answering machines are eligible, while those resulting repeatedly in busy signals are not eligible. The limitation of this method is that the basis in which the disposition codes are allocated

[9] Some of the literature on the response rate refers to this method as CASRO type II as it is proposed by the American Survey Research Organisations.

may not solely determine the eligibility. For example, a ring-no-answer alone is not enough to identify a case as being ineligible.

As for the business offices[10] approach, survey researchers sometimes contact local telephone business offices to enquire about the status of the unknown numbers (Frankel *et al*; 2003). However, this method is considered to be both money and time consuming, in addition to the fact that business offices usually refuse to give out information about phone numbers.

Many studies have applied the above methods to estimate the eligibility rate among the cases of unknown eligibility. For example, Barron, Khare and Zhao (2008) applied the proportional allocation approach to estimate 'e', to calculate the response rate for the National Immunization Survey's Cell Telephone Pilot study (NIS-CTP). Gasteiz (2007) indicates that the minimum and maximum allocation method was used ('e' was assumed to be 100%) to estimate the eligible cases among those cases where eligibility is unknown in the Population in Relation to Activity Survey (PRA). In a list-assisted RDD telephone survey about adolescent substance abuse, the Survey and Evaluation Research Laboratory (SERL) applied the proportional allocation approach to estimate the response rate (Ellis, 2000).

However, each of the methods used has its limitations, and as Smith (2003) states "*At present none can be considered a gold standard for calculating "e"*". In addition, most of these methods have mainly been implemented in cross-sectional studies. Longitudinal

[10]A telephone business office is a special type of firm in the United States that possesses telephone numbers directories for specific geographical areas. Such companies provide telephone numbers, addresses and directions for businesses and individuals within the specific areas.

surveys, on the other hand, may benefit from the investigation of an alternative method that takes the longitudinal aspect of eligibility into account.

2.5 The effect of unknown eligibility on response rate, contact rate and co-operation rate

When the data collection stage is completed, survey organizations usually publish some statistics such as the response rate, contact rate and co-operation rate, to reflect the main features of the data and inform data users about the quality of the data that the survey has gathered. These rates are widely reported in research reports. They indicate the quality of the survey and the effort put forward to contact sample members and achieve interviews. Also, these rates are, sometimes, used to compare survey quality between surveys, survey organisations and countries. In addition, response rates are important indicators of the likelihood of non-response bias. On the absence of good estimates of the differences between respondents and non-respondents in terms of the measurements used to construct the estimate in question, low response rates may be taken as indicators of a potential for bias.

However, unknown eligibility can affect the calculation of these rates because each of them is defined as a ratio that contains the number of eligible sampled units in the denominator. Thus, the identification of eligible cases will affect the base on which these rates are calculated. Consequently, an incorrect classification of cases whose eligibility is unknown will result in under or over estimating response, contact and co-operation rates.

Response rate

The response rate measures the percentage of the completed interviews out of all the eligible units (CASRO, 1982; AAPOR, 2004). From this definition, it is obvious that the calculation of response rate depends on a precise identification of outcome codes for sample cases. For example, does a partial interview count as 'completed interview'? Also, in a random dialling digit (RDD) survey, can a 'ring-no-answer' indicate eligibility even though contact is not successfully established?

Different survey designs may result in different outcome codes (or different distribution of cases over the same codes if the studied populations are different), and hence different calculations of response rates. Thus, it is not always possible to compare response rates between surveys because of design differences or population differences. In recent decades, efforts have been made by different organisations to standardise the calculation of response rates. For example, in the USA, the American Association for Public Opinion Research (AAPOR) published standard definitions for final disposition of case codes for both RDD and in-person surveys (AAPOR, 1998). Also, in the UK, Social Survey Division of ONS and the National Centre for Social Research developed a proposal for outcome codes and response rate calculation for use with government social surveys (Beerten, Lynn, Laiho and Martin, 2000).

In any case, ineligible sample units should not be included in the calculation of the response rate if the rate is to be computed accurately. Based on the definition of the response rate, a general formula for calculating the rate can be written as

$$RR = \frac{\text{Number of eligible sample units with completed interviews}}{\text{Number of eligible sample units}} \tag{2.1}$$

Where RR is response rate. However, it is impossible to calculate the denominator precisely, if eligibility is unknown for some of the sample members. In almost every survey, as long as there is an incidence of non-response, there will be a number of cases whose eligibility remains unknown.

Thus, estimating the Number of Eligible Units whose Eligibility is Unknown (NEUEU) is a crucial component of the calculation of the response rate (Alsnih and Stopher, 2004). If the NEUEU is estimated, the response rate can be calculated using (2.2).

$$RR = \frac{\text{Number of eligible sample units with completed interviews}}{\text{Number of eligible units whose eligibility is known} + \text{NEUEU}} \tag{2.2}$$

Overestimating NEUEU leads to underestimating the response rate, while underestimating NEUEU results in overestimating the response rate. Therefore, regardless of the method used to estimate NEUEU, it is advisable to utilize a value of NEUEU that does not inflate the response rate and hence give a false sense of valuing the quality of the data.

Contact rate

The contact rate (CR) denotes the proportion of sample members who were successfully contacted, even if they refused to participate in the survey or were unable to provide any type of information (Gasteiz, 2007). A general formula for calculating CR is given by equation (2.3).

$$CR = \frac{\text{Number of eligible sample units in which contact was made}}{\text{Number of eligible sample units}} \tag{2.3}$$

Similar to the response rate, the calculation of contact rate requires the estimation of NEUEU after which the rate can be calculated using (2.4).

$$CR = \frac{\text{Number of eligible sample units in which contact was made}}{\text{Number of eligible units of known eligibility} + \text{NEUEU}} \qquad (2.4)$$

Co-operation rate

The co-operation rate (CoR) measures the proportion of achieved interviews among the cases in which contact was made (Gasteiz, 2007). CoR can be calculated using equation (2.5).

$$CoR = \frac{\text{Number of sample units in which interview was conducted}}{\text{Number of eligible sample units in which contact was made}} \qquad (2.5)$$

In many social science surveys (including the BHPS), a sample member is eligible if they are alive. In this case, calculating the co-operation rate does not require an estimate of NEUEU. This is because the denominator in the co-operation rate only consists of sample members who are successfully contacted and hence eligible (i.e. alive and living in the geographical area covered by the survey).

Survey researchers recommend the calculation of contact and co-operation rates alongside the response rate (Beerten, Lynn, Laiho and Martin, 2000). The response rate on its own tells us little about the mechanism underlying the non-response process. By quoting the response rate alone, it is not possible to reflect on refusal and non-contact, for example. In turn, calculating contact and co-operation rates helps analysts to understand whether refusal or non-contact is the major component of non-response in the survey, and also helps survey organisations to tackle the problem more appropriately by dealing with the different causes separately.

The response rate may be redefined in terms of the contact rate and the co-operation rate as:

Response rate (RR) = Contact rate (CR) * Cooperation rate (CoP) (2.6)

In short, with unknown eligibility, the calculation of the different types of outcome rate encounters the problem of accurate estimation of the number of eligible sample units whose eligibility is unknown (NEUEU). NEUEU can be estimated using a number of practical methods (reviewed in section 2.4). Thus, regarding the calculation of the survey quality measures, unknown eligibility can –to an extent- disturb the calculation of some of these measures. Though, with a good estimation of NEUEU, one can still calculate the quality measures to the best possible approximation. However, with weighting, unknown eligibility may be more problematic. Our substantive investigation in this chapter is centred around this issue. In particular, we study the negative impact of unknown eligibility on the SWA and we suggest alternatives.

2.6 Unknown eligibility and the standard weighting approach (SWA)

In longitudinal surveys, apart from disturbing the calculation of the quality measures, unknown eligibility can also distort non-response weighting. Non-response weighting assigns higher values to some of the eligible respondents in the survey, in order to increase their influence so as to represent eligible individuals who are missing due to non-response (Biemer and Christ, 2008; Lynn, 2005). Therefore, in order for the weights to modify the sample correctly, they should be estimated using eligible sample members only. Including ineligible sample members in the non-response model that is used to derive the weights may lead to incorrect modification of the sample as will be demonstrated next.

During weights creation, classification of sample cases as either eligible or ineligible must be made for every sample member. Clearly, unknown eligibility is a predicament in this situation. As we discussed in the introduction of the thesis, and the chapter, the SWA attempts to overcome this difficulty by assuming that all sample members whose eligibility is unknown are eligible. However, if some of these cases are actually ineligible, the weights may not be calculated accurately. More importantly, if a large number of the ineligible cases (who are not known as such) are clustered within specific weighting classes, the weights may result in estimates that are biased towards characteristics of sample members from these classes.

To illustrate this problem for a given survey that applies the SWA, suppose that

n_{ijk} indicates the number of sample units (where *i*, *j* and *k* denote eligibility status, knowledge of eligibility, and survey response status respectively); and that

i = 1 if eligible; 2 if ineligible (actual status, regardless of whether this is known);

j = 1 if eligibility status is known; 2 if it is not known;

k = 1 if survey respondent; 2 if non-respondent.

Thus,

$\sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} n_{ijk} = n_{\bullet\bullet\bullet}$ is the total sample size.

We can assume that $n_{121} = n_{211} = n_{221} = 0$ (i.e. that all respondents are eligible and known to be eligible).

Accordingly,

n_{111} is the number of respondents;

n_{112} is the number of non-respondents known to be eligible;

n_{212} is the number of non-respondents known to be ineligible; and

$n_{122}+n_{222}$ is the number of non-respondents whose eligibility is unknown (where n_{122} is the number of non-respondents whose eligibility is unknown but are actually eligible, and n_{222} is the number of non-respondents whose eligibility is unknown but are ineligible).

For illustration purposes, let us assume that the weighting is done using weighting classes method. For a given weighting class 'c', the response probability ($Þ_c$) is calculated by dividing the number of responding units by the total number of eligible sample members in the class; and the relevant non-response adjustment weight (w_c) is calculated as the inverse of the response probability in the class.

Now, with unknown eligibility, the SWA assumes that sample members whose eligibility is unknown ($n_{122}+ n_{222}$) are eligible. Thus:

$$Þ_{1c} = \frac{n_{111}}{n_{11.}+ n_{122}+ n_{222}} \quad \text{resulting in } w_{1c} = \frac{n_{11.}+ n_{122}+ n_{222}}{n_{111}}$$

But, if we had perfect information about eligibility, this would be:

$$Þ_{2c} = \frac{n_{111}}{n_{11.} + n_{122}} \quad \text{resulting in } w_{2c} = \frac{n_{11.} + n_{122}}{n_{111}}$$

Note that w_{1c}>w_{2c} as a consequence of not being able to correctly identify eligibility for sample units in n_{222} (ineligible units amongst units of unknown eligibility) during the calculation of w_{1c}. Ideally, the SWA should exclude sample units in n_{222}, and should produce w_{2c}. However, because the SWA assumes that all units of unknown eligibility (n_{122}+ n_{222}) are eligible, it results in w_{1c}. In turn, w_{1c} will mistakenly over weight the responding cases in class 'c' if n_{222} is large. In addition, over time, the size of w_{1c} may continue to increase incorrectly as more cases are added to n_{222}. In other words, as more waves are conducted, the relevant weight in class 'c' will increase (mistakenly), if some of the new non-responders in the class become ineligible but are not identified as such. As a result, the increase in weights size may increase the standard errors of weighted estimates leading to less precision and less statistical power.

Moreover, and even more importantly, this might cause a larger problem if ineligible cases among cases of unknown eligibility are not evenly distributed across weighting classes. If, in this case, the weighting classes are associated with some of the survey key variables, respondents in classes with larger proportions of unidentified ineligible cases, will have greater contribution (with the larger weights) to estimates constructed using the variables in question, but by using influence from units that are not part of the population of interest (n_{222}). Consequently, the weighted estimates will be biased towards the characteristics of respondents from classes where more ineligible cases are not identified.

Obviously, the problem also applies if a model-based method is used for weighting. In this case, units in n_{222} who possess a given set of characteristics that determine the relevant probability of response, may falsely induce the model to estimate a smaller response probability (as they are non-responders) and hence generate larger weights for the responding cases who share the same set of characteristics that are used as predictor variables in the model. Had the cases in n_{222} been identified and dropped from the weighting model, the estimated probability of response which is determined by the set of characteristics that cases in n_{222} possess would increase in size resulting in a smaller weight for the responding cases who share these characteristics.

For most longitudinal surveys that intend to study a resident population in a given geographical area, death is an obvious ineligibility criterion. Accordingly, if the SWA is applied, deceased sample members who are not reported as such will be assumed as alive non-responders, and will be used in the weights calculation as was just explained. However, health studies have shown that death is associated with socio-demographic characteristics such as age and gender (Singh-Manoux *et al*, 2008; Dr Foster, 2004). That is to say, in most parts of the world women are expected to live longer than men, and mortality rates are higher among older age groups than among their younger counterparts. Therefore, estimates constructed based on a SWA which might mistakenly include dead (ineligible) sample members in the weights' calculation may be biased towards characteristics of respondents from classes with higher death rates (e.g. respondents in the older age groups). Thus, the treatment of unknown eligibility cases by the SWA may not be the best option if other alternatives are possible.

2.7 The alternative weighting approach (AWA)

As mentioned previously, the most common forms of ineligibility are: death, moving from the geographical area covered by the survey and being institutionalised. As is set out in the introduction of this chapter, the investigation here will be limited to death.

The SWA results in relatively larger weights for the responding cases whose characteristics are similar to the deceased sample members. As a counter effect, the AWA will be based on a mechanism (as will be shown) that reduces the weights of these responding cases by implementing an adjustment that relies on the population mortality rates to estimate the adjustment factor. The rationale for estimating and applying this adjustment factor is explained in what follows.

The sample – if representative - is a smaller image of the population. Thus, the rate of a given phenomenon in the population should be equal to the rate of the same phenomenon in the sample, under expectation. Based on this logic, the AWA involves estimating expected sample survival proportions, for classes identified by age and gender, using the population mortality data from the external source. It also calculates the equivalent observed rates in the sample. The latter would be obtained by assuming that those whose eligibility is unknown are alive. Note if those with unknown eligibility status in the sample are assumed to be alive, the observed survival rates in classes with larger proportions of unknown eligibility cases are likely to be higher in the sample than in the equivalent classes in the population. This is simply because some of the unknown eligibility cases within these classes in the sample may not be alive. Also, the standard weights in such classes are large (incorrectly). However, the ratio of the survival rate in the population to the survival rate in the sample in each class, which will range between 0 and 1, can be used to reduce the relevant weights to balance eligibility.

For example, assume that the survival rate in a given population is 80%; and that a representative sample from this population shows that 50% are known to be alive, 10% known to be dead and 40% are of unknown survival status. If the weighting assumes that those with unknown survival status are alive, the weights will increase the values of the responding sample to represent 90% survival (50% + 40%) while it should only represent 80% (the true rate in the population). Thus, it would be appropriate to adjust (multiply) the weights by 8/9 (80%/90%) to correct the survival imbalance. In this example, 8/9 is the adjustment factor. It is calculated as the ratio of the survival rate in the population to apparent survival rate in the sample.

To show the calculation of the adjustment factor in terms of notations let $Þ_S$ and $Þ_P$ denote the survival rate in the sample and in the population respectively. Note that $Þ_S$ contains cases that are assumed to be alive (i.e. cases with unknown eligibility status). Thus:

$$Þ_S \geq Þ_P \tag{2.7}$$

if

$$ad{*}Þ_S = Þ_P \tag{2.8}$$

Then *ad* is a fraction.

ad is the adjustment factor that equalises the survival rates in the sample and the population. From equation (2.8)

ad= $Þ_P/Þ_S$ (2.9)

From equation (2.9) it can be notice that *ad* takes a value between 0 and 1 since its denominator is larger.

Our AWA in this chapter relies on the *ad.* After creating the weights using the SWA, the AWA, will classify the sample members in classes identified by age and gender. For each class, the AWA will calculate the *ad* as in equation (2.9). The standard weight of a given respondent will be multiplied by the value of *ad* in the class to which the respondent belongs. In classes where all units are known to be alive ($Þ_P \approx Þ_S$), the value of the *ad* will be 1 resulting in no effect on the relevant weights. In classes where more units are assumed to be alive but they are actually deceased ($Þ_P < Þ_S$), the value of the *ad* will take a value below one, and will therefore decrease the weights values of the responding cases in these classes reducing the negative effect of unknown eligibility on weighting.

At first sight, the AWA may seem as if it is merely a standard post-stratification method. Partly because both the AWA and post-stratification classify the sample into classes for which information is known for the same classes in the population; and partly because the weights in these classes are adjusted further by a constant that is constructed based on the population information. Although the two methods share similarities, they aim at, and achieve, different results. Differences between our AWA and post-stratification can be explained in a number of dimensions:

First, the targets in post-stratification (often population totals) are known for the defined classes, whereas in the AWA the targets (eligibility rates) need to be estimated based on a combination of sample information (observed survival rates) and external information (mortality rates). Second, the targets in post-stratification are estimates of the population distribution; whilst in the AWA, the eligibility rates calculated from the population information are estimates of the distribution of the sample (true rates of eligibility). Third, post-stratification should work in an upwards direction (increases the weights) to represent eligible units in the population that are missing from the sample, while the AWA only works in a downwards direction (it decreases the weights since $0 < ad < 1$) to reduce influence of ineligible units that were assumed to be eligible sample units. Fourth, the main goal of post-stratification is to adjust for non-coverage and sampling error, where the aim of the AWA is to correct for estimation error that would otherwise be caused by an incorrect assumption underpinning the non-response adjustment weighting.

A second alternative approach

An interesting alternative with respect to determining the eligibility status for those whose eligibility is unknown may involve a case-level survival/death imputation. In this approach, one can look at 'eligibility status' as a variable that indicates whether a sample member is eligible, ineligible or of unknown status. Within the variable 'eligibility status' values indicating unknown eligibility status can be treated as missing values. Imputation can then be used to impute the missing values as either eligible or ineligible (alive or dead). As a result, eligibility status will be decided, at the case level, for all sample members. Accordingly, the alternative weighting will exclude ineligible (known+imputed) sample members from the weights creation. In this approach, variables such as gender, age, race and health indicators may be strong candidate variables as predictors in the imputation model because of their strong relationship with mortality.

In this chapter, although our focus is on the adjustment factor approach that was previously introduced and labelled as the AWA, we also apply the imputation of unknown eligibility status. We create weights based on the outcome of the imputation of eligibility status as a second alternative weighting. The main purpose of the second alternative weighting is to test the sensitivity of the result from the AWA. Details of the imputation of eligibility status and the resultant set of weights are given in the next section.

The advantage of the second alternative approach (imputation-based) is that it is relatively easy to apply since it uses information from within the survey (no need for seeking and utilising information from an external source). Thus, it might seem more practical to implement this approach as opposed to the AWA. However, there are some limitations to the second alternative approach that may raise concerns about the imputed eligibility status.

First, to accurately impute the missing values of ‘eligibility status’ as either eligible or ineligible, the missingness mechanism should be MAR. If the MAR is not satisfied, the missingness mechanism may still depend on the missing values even after controlling for the observed variables used in the imputation. However, MAR is not an assumption that can be supported by the nature of missingness in the variable ‘eligibility status’ (recall that a missing value here indicates that eligibility is unknown). Although some values can be missing for random reasons that are not correlated with eligibility status, it is likely that many values are missing because their corresponding sample members are ineligible (deceased). In the latter case, there is a direct relationship between the missingness mechanism and the variable of interest (eligibility status). If the missingness still depends on the true eligibility status (death) after controlling for the observed variables the MAR assumption is violated.

Second, it could be tricky to take into account the time dependency of the predictor variables used in the imputation. For example, to impute eligibility for the BHPS sample members whose eligibility is unknown, one should use predictors from the same wave for both the cases used in the imputation and the cases whose eligibility is to be imputed. Namely, predictor variables should be from the wave at which unknown eligibility cases were last observed. However, the waves at which sample members of unknown eligibility were last observed are different which makes it challenging to incorporate the time dependency. Furthermore, the predictor variables that are suggested for the imputation may not be available across all waves.

Despite a few limitations, the second alternative approach may still offer a better strategy to dealing with the cases of unknown eligibility whilst weighting in comparison with the SWA. This is because the second alternative approach uses information from the multivariate structure of the data (the relationship between eligibility status and the variables used in the imputation), which the SWA largely ignores.

2.8 Methodology

For our analysis we used respondents of the original sample of the BHPS who were aged 16+ at the start of the survey (in 1991). These are 10,248 sample members who responded at wave 1 (and were therefore all alive at that time). By the end of wave 18 (in 2008), eligibility was known for 69.6% (7,130) of these sample members. The remaining 30.4% (3,118) were of unknown eligibility status. For those whose eligibility is known by wave

18, 5,588 (78.4%) of them are known to be alive (eligible), whereas 1,542 (21.6%) are known to be ineligible.

We used this sample to investigate the issue of unknown eligibility and weighting by wave 18. The investigation involved creating non-response longitudinal weights at wave 18 based on the SWA, the AWA and the second alternative approach (imputation-based method) as well as conducting substantive analysis using the resultant sets of weights. For our substantive analysis we used a balanced panel of those who responded in all of the 18 waves. These are 40% (4,097 respondents) of the original sample.

We classified the cases of the original sample of the BHPS by gender and their single-year age in wave 1[11], and implemented the following steps:

- estimating the survival rates in the sample in 2008 (wave 18) for each class of gender and age, by assuming that all observed statuses are correct and that all those whose survival status is unknown in 2008 are alive;
- using annual population mortality statistics for each year between 1992 and 2008 from two sources: Office for National Statistics (ONS) and government statistics (available on statistics.gov.uk) to construct the expected 1991-2008 survival rates for each class defined by gender and single year of age in 1991 (wave 1);
- calculating the adjustment factor for each class of gender and age by dividing the population survival rate by the sample survival rate;
- creating the standard longitudinal weights (*SWs*) at wave 18 using the SWA;

[11]It is necessary to use age at wave 1 to be able to estimate survival rates by 2008 from population information as will be shown later.

- creating the main set of the alternative longitudinal weights at wave 18 (AWs_1) using the AWA (i.e. multiplying the standard weights in each class of gender and age by the relevant adjustment factor);
- carrying out eligibility status imputation for sample members whose eligibility is unknown, and creating a second set of alternative longitudinal weights (AWs_2) by excluding sample members whose eligibility status is imputed as dead prior to applying the SWA (for sensitivity check purposes); and
- assessing the effect of the AWA by conducting weighted analysis on subjective health status using a balanced panel from wave 1 to 18 with SWs, AWs_1 and AWs_2 separately and compare the results.

The details of each of these steps are provided in what follows.

2.8.1 Calculating survival rates in the sample

The BHPS provides details about the contact outcome at every wave. The main reported outcomes are full interview, proxy interview, telephone interview, refusal, non-contact, in institution, out of scope or dead. In a given wave, these outcomes lead to three categories of sample members in terms of eligibility status: a) *Eligible sample members (alive) whose eligibility status is known* are those who gave full interview, proxy interview, telephone interview or refused to participate; b) *Ineligible sample members whose eligibility status is known* are those who are in institution, out of scope or dead; and c) *Sample members whose eligibility is unknown* are those who were not contacted. Based on this classification, survival rates in the sample could be calculated by assuming that those whose eligibility is unknown are eligible. Thus, the numerator for survival rates consists of sample members from category (a) and (c).

Accordingly, for our original sample, which was classified by gender and single-year age, we first calculated the rates of those who were known to still be eligible at wave 18 (i.e. in 2008) for every class. These rates are presented separately in table 2.1. In addition, for the same classes of gender and age, we calculated the rates of those whose eligibility is unknown by wave 18 and also presented these separately in table 2.1. The combination (sum) of the known eligibility and unknown eligibility rates are the survival rates in the sample. However, for the purpose of understanding the distribution of the sample in terms of known/unknown eligibility status, these rates are presented separately in table 2.1. As will be shown in the next section, these are combined as a single-rate (survival rates in the sample) and presented together (in table 2.2) with the survival rates from the population. As for those who were known to be ineligible subsequent to wave 1 (category b), these were also calculated and displayed in table 2.1.

Looking at the rates in table 2.1, overall, 46.4% of the males in the original sample of the BHPS who responded in the first wave (1991) are still eligible in 2008 while 20.1% of them are known to be ineligible and the remaining 33.5% are of unknown eligibility. As for females, the corresponding rates are 50.6%, 19.3% and 30.1% respectively. This shows that the known eligibility rate for females is higher than for males, whereas unknown eligibility is higher among males than among females. Perhaps this is because females are more likely to respond than males, but also may be because females are rather easier to track over time than males (e.g. in comparison with males, it is more likely for females to reside with children which in turn makes tracking easier).

In general, for both gender types, the eligibility rate increases with age until it reaches its peak at the age of 38 for males (67.4%) and at age 30 for females (70.2%). Then it starts declining as age increases, to reach its nadir (0%) at the ages of 81 and 82 for males and females respectively. Also, the ineligibility rate increases with age, more so in older ages than in younger ages. However, it increases faster for males than females. Ineligibility rates of 80% or above were registered for males as early as age 77, whereas for females rates as high as 80% were not registered until age 83. These findings are consistent with the literature on mortality, indicating that death rates are always higher amongst older age groups (Singh-Manoux *et al*, 2008) and life expectancy among females is higher than among males (Dr Foster, 2004).

Additionally, for both males and females, the unknown eligibility rates are higher with younger sample members than with their older counterparts. The explanation for this is that young individuals are more mobile and hence more difficult to contact and follow overtime compared to older sample members. However, for both males and females, some of the sample members who started the survey at older ages (80+) registered considerable unknown eligibility rates by wave 18. For example, 33.3% of men and 22.2% of women who were aged 89 in wave 1 are of unknown eligibility at wave 18 (17 years later). It might be plausible to assume that some of these cases are ineligible (deceased). Nevertheless, the SWA, would assume these cases to be eligible.

Table 2.1. Proportions of eligible, unknown eligibility and ineligible cases in 2008 for the original sample of the BHPS (1991) by gender and single-year age.

	Eligible		UE		Ineligible	
Age	Male	Female	Male	Female	Male	Female
16	58.6	59.5	39.8	39.2	1.6	1.2
17	56.0	62.7	42.1	35.9	1.9	1.4
18	54.1	55.6	43.5	42.9	2.4	1.5
19	52.2	50.3	45.3	48.1	2.6	1.6
20	46.7	52.0	50.6	46.4	2.7	1.6
21	56.9	62.9	40.4	35.3	2.7	1.8
22	47.1	58.4	49.8	39.7	3.2	1.9
23	48.0	60.9	48.9	37.2	3.1	2.0
24	46.9	66.9	49.7	31.0	3.4	2.1
25	44.3	51.1	52.0	47.0	3.7	1.9
26	54.3	61.1	42.0	36.8	3.8	2.1
27	46.0	61.6	50.0	36.2	3.9	2.2
28	52.7	56.1	43.4	41.7	3.9	2.2
29	57.1	53.6	38.8	44.1	4.1	2.3
30	52.2	70.2	43.6	27.5	4.3	2.3
31	52.8	62.4	42.8	35.2	4.4	2.4
32	52.9	58.6	42.7	39.0	4.5	2.4
33	61.1	59.8	34.3	37.6	4.6	2.6
34	53.5	67.7	41.9	29.8	4.6	2.6
35	58.5	64.9	36.9	32.4	4.7	2.6
36	49.0	56.0	46.3	41.3	4.7	2.7
37	61.6	65.6	34.4	31.6	4.0	2.8
38	67.4	59.6	27.8	37.5	4.8	3.0
39	60.8	65.7	36.3	31.3	2.9	3.0
40	55.5	58.3	39.7	38.2	4.9	3.5
41	53.5	60.8	40.7	35.4	5.8	3.8
42	54.0	55.8	40.1	40.3	5.9	3.9
43	53.9	63.6	39.2	32.2	6.8	4.2
44	59.2	59.7	33.9	35.6	6.9	4.8
45	56.3	64.7	35.8	30.2	7.9	5.2
46	44.6	66.7	47.4	27.8	8.0	5.5
47	49.4	55.7	42.5	38.4	8.1	5.9
48	55.4	50.0	36.1	43.2	8.5	6.8
49	61.9	63.2	28.0	29.9	10.1	7.0
50	56.9	62.7	30.6	29.7	12.4	7.6
51	46.8	46.9	39.9	44.2	13.4	8.9
52	50.6	63.2	37.0	27.6	12.4	9.2
53	60.3	55.1	28.2	34.8	11.5	10.1
54	50.0	56.7	29.8	32.2	20.2	11.1
55	46.3	59.7	32.4	31.6	21.3	8.8
56	60.4	57.1	25.9	32.5	13.8	10.4
57	41.5	57.8	34.8	32.8	23.7	9.4
58	58.3	59.3	18.8	27.1	22.9	13.8
59	43.9	58.4	28.3	27.2	27.8	14.4
60	58.5	57.8	26.5	27.0	26.4	15.2
61	44.2	43.9	25.0	37.6	30.8	18.5
62	41.3	38.2	30.7	40.9	28.0	20.9
63	42.4	43.5	20.6	34.0	37.0	22.5
64	33.9	39.6	23.9	34.5	42.1	25.9
65	34.6	38.2	23.1	31.1	42.3	30.7
66	22.8	38.8	27.1	27.8	50.1	33.5
67	34.6	35.6	21.8	27.4	43.6	37.0
68	30.4	31.1	17.9	28.7	51.8	40.2
69	14.8	27.1	31.5	33.9	53.7	39.0

* The table is continued in the next page. Entries are percentages per 100 persons. UE refers to unknown eligibility.

Table 2.1 (continued)

	Eligible		UE		Ineligible	
Age	Male	Female	Male	Female	Male	Female
70	25.4	26.3	23.7	24.1	50.9	49.6
71	10.8	16.9	24.6	29.1	64.6	54.0
72	16.7	12.7	19.4	32.1	63.9	55.2
73	6.1	19.4	22.1	27.8	71.8	52.8
74	8.8	12.3	17.7	29.8	73.5	57.9
75	5.4	12.8	29.7	25.5	64.9	61.7
76	10.3	15.0	13.8	27.5	75.9	57.5
77	4.2	7.7	8.3	25.0	87.5	67.3
78	3.7	2.0	14.8	28.0	81.5	70.0
79	0.0	2.6	17.1	28.2	82.9	69.2
80	5.6	0.0	33.3	21.9	61.1	78.1
81	0.0	1.9	5.9	26.4	94.1	71.7
82	0.0	0.0	17.4	27.6	82.6	72.4
83	0.0	0.0	5.1	20.0	94.9	80.0
84	0.0	0.0	10.0	8.3	90.0	91.8
85	0.0	0.0	20.0	21.4	80.0	78.6
86	0.0	0.0	8.8	6.3	91.2	93.8
87	0.0	0.0	8.8	11.1	91.2	88.9
88	0.0	0.0	8.2	33.3	91.8	66.7
89	0.0	0.0	33.3	22.2	66.7	77.8
90 or over	0.0	0.0	8.4	8.3	91.6	91.7
Overall	**46.4**	**50.6**	**33.5**	**30.1**	**20.1**	**19.3**

*Entries are percentages per 100 persons. UE refers to unknown eligibility.

2.8.2 Estimating survival rates from population information

To estimate the survival rates from population information, ideally, we would want to have mortality rates by gender and single-year age for the population in England, Wales and Scotland (the BHPS population) for every year from 1992 to 2008 since we know that everyone in the sample was alive in 1991 (wave 1). If this information were available, estimating survival rates by 2008 from population information would be fairly simple[12]. However, our analysis here encountered two obstacles in this regard:

[12]One may first turn the annual mortality rates into survival rates. Then, starting with rates in 1992, the survival rate by 2008 for every age is the product of the survival rates for the consecutive ages in the successive years up until 2008. An example of this is given later during the estimation of survival rates from population information.

First, mortality statistics are not typically available for the population in England, Wales and Scotland together (to match the target population of the BHPS original sample). Instead, most of the available informative statistics on mortality from the Office for National Statistics (ONS) or other government statistics authorities in the UK are for the population in England and Wales. Thus, to be able to use the available statistics, our calculation here assumed that survival/mortality rates for the population in England and Wales are the same as those for the population in England, Wales and Scotland.

Second, although the ONS releases mortality rates for the population in England and Wales by gender and single-year age, unfortunately these rates are not available for all years in the period of 1992-2008. For years before 1999, the rates were released by gender and 10-year age bands.

However, ONS releases the number of registered deaths for the population in England and Wales by gender and single year of age for the years in question. Appendix B.1 and B.2 show these numbers for all years from 1992 to 2008 by single-year age and for males and female respectively. Also, official government statistics websites (statistics.gov.uk AND nationalarchive.gov.uk) publish the number of residents for the population in England and Wales by gender and single year of age for every year in the period of time 1992-2008. These numbers are shown for all years by single-year age and for males and females in appendix B.3 and B.4 respectively. We used this information (number of deaths from ONS and number of residents from statistics.gov.uk) to calculate the mortality rates by gender and single-year age for every year from 1992 to 2008.

For every year, and for every class of gender and single-year age, we calculated the mortality rate by dividing the number of registered deaths in the year by the total number of residents in the same year[13]. We used the resultant mortality rates to estimate the survival rates by 2008. This, first, involved calculating the annual survival rates for all years from 1992 up to 2008 (by subtracting mortality rates in these years from 1). Then, starting with the rates in 1992, the 1991-2008 survival rate for every single year of age, for each gender, was calculated as the product of the rates for the consecutive ages in the successive years up until the rate of the corresponding age in 2008. For example, for males aged 16 in 1991 (wave 1), the 1991-2008 survival rate was estimated by multiplying the survival rate of age 17 (in 1992) by the survival rate of age 18 (in 1993) and so forth up to the survival rate of age 33 (in 2008). This was done separately for both gender types and for all single-year ages. All the calculated survival rates by 2008 are shown by gender and age with their corresponding rates from the sample, and the relevant adjustment factors in table (2.2).

As can be seen from table 2.2, for both gender types and all ages, the survival rates calculated from the sample are larger than those estimated from the population information (recall that sample survival rates here include those whose eligibility is unknown). This confirms our hypothesis that some of the unknown eligibility cases might not be alive (eligible). Nevertheless, overall, the differences are not worryingly large, and suggest only a small number of deaths amongst sample members whose eligibility is unknown. In fact, for all ages under 60, survival rates in the sample and the population are

[13] This is how the ONS calculates mortality rates by gender and single-year age every year from 1999 onwards (see for example National Statistics, 1999).

very similar indicating that most sample members who started the survey aged 59 or under, and whose eligibility status is unknown by 2008 (wave 18) may still be alive.

In turn, the differences between survival rates in the sample and the population start increasing, for both gender types, from age 60 onwards. Older ages (70+) registered even larger differences compared to the relatively younger ones (60-70). These results indicate that a considerable proportion of those who started the survey aged 60+, and whose eligibility is uncertain by the time of wave 18 might be deceased. This is particularly more likely for those aged 70+ at wave 1 than for those aged between 60-70.

In addition, these results suggest that sample members whose eligibility is unknown but are actually ineligible are not evenly distributed across the sample. Instead, these cases are rather clustered at one end of the age spectrum (age 60+). Consequently, the SWA which assumes that these cases are eligible, will mistakenly increase the size of the weights values for respondents aged 60+ as was explained earlier. As a result, respondents aged 60+ will have more influence (incorrectly) on weighted estimates, and therefore the results may be less precise, but might also be biased towards characteristics of those aged 60 or over. This is especially more likely in substantive analyses on social phenomena that are directly correlated with age such as changes in health status (the outcome variable in our substantive analysis).

2.8.3 The adjustment factor (*ad*)

The survival rates from the sample and the population in table 2.2 were used to calculate the *ad* (presented in the same table) by gender and age. For each class of gender and age, the *ad* was calculated as the ratio of the survival rate in the population to the survival rate in the sample as was shown in equation 2.9.

Figure 2.1 presents the distribution of the *ad* by gender and age in a two-dimension graph. As can be noticed from the graph, for both men and women who began the BHPS aged between 16 and 59, the values of the *ad* are almost the same and approximately equal to 1. These values of *ad* were expected since there were no large differences between the survival rates in the sample and the population for ages under 60. However, for both men and women aged 60+, the factor shows a drastic decrease in its values as age increases because of the larger differences between the sample and the population survival rates of those aged 60 or above.

Since the *ad* will be used in the AWA to adjust the weights resulting from the SWA, the results here suggest that the AWA may not have much effect on weights values of the set of responding sample who started the survey aged between 16 and 59 ($0.91 \leq ad \leq 0.99$). In contrast, the AWA will have a larger effect (reduction) on the weights of those started the BHPS aged 60+ ($0.03 \leq ad \leq 0.86$). Thus, while the SWA may incorrectly increase the sizes of the standard weights of those aged 60+ by including a proportion of ineligible cases in the weighting, the AWA decreases the values of these weights using our proposed method of adjustment. As a result, the negative effect of unknown eligibility on weighting will be reduced and hence estimates may be more precise and less biased.

Table 2.2 Estimated survival rates by 2008 for the BHPS original sample from population information, assumed survival rates in the sample and the adjustment factor.

	Male			Female		
Age	Population	Sample	*ad*	Population	Sample	*ad*
16	97.6	98.4	0.99	98.3	98.7	0.99
17	97.5	98.1	0.99	98.2	98.6	0.99
18	97.4	97.6	0.99	98.2	98.5	0.99
19	97.4	97.5	0.99	98.1	98.4	0.99
20	97.2	97.3	0.99	98.0	98.4	0.99
21	97.0	97.3	0.99	97.9	98.2	0.99
22	96.8	96.9	0.99	97.9	98.1	0.99
23	96.7	96.9	0.99	97.8	98.1	0.99
24	96.5	96.6	0.99	97.7	97.9	0.99
25	96.2	96.3	0.99	97.7	98.1	0.99
26	96.0	96.3	0.99	97.6	97.9	0.99
27	95.9	96.0	0.99	97.5	97.8	0.99
28	95.9	96.1	0.99	97.5	97.8	0.99
29	95.7	95.9	0.99	97.4	97.7	0.99
30	95.6	95.8	0.99	97.4	97.7	0.99
31	95.6	95.6	0.99	97.3	97.6	0.99
32	95.5	95.6	0.99	97.3	97.6	0.99
33	95.4	95.4	0.99	97.2	97.4	0.99
34	95.3	95.4	0.99	97.2	97.5	0.99
35	95.2	95.4	0.99	97.1	97.3	0.99
36	95.1	95.3	0.99	97.1	97.3	0.99
37	95.0	96.0	0.98	97.0	97.2	0.99
38	94.7	95.2	0.99	96.4	97.1	0.99
39	94.6	97.1	0.97	96.3	97.0	0.99
40	94.3	95.2	0.99	96.2	96.5	0.99
41	93.8	94.2	0.99	95.8	96.2	0.99
42	93.1	94.1	0.98	95.4	96.1	0.99
43	92.5	93.1	0.99	95.0	95.8	0.99
44	91.8	93.1	0.98	94.5	95.3	0.99
45	90.9	92.1	0.98	94.0	94.9	0.99
46	90.4	92.0	0.98	93.6	94.5	0.99
47	89.2	91.9	0.97	93.0	94.1	0.99
48	88.2	91.5	0.96	92.3	93.2	0.99
49	87.1	89.9	0.96	91.7	93.1	0.98
50	85.6	87.5	0.97	91.0	92.4	0.98
51	84.5	86.7	0.97	90.0	91.1	0.98
52	82.5	87.6	0.91	88.9	90.8	0.97
53	80.8	88.5	0.95	87.8	89.9	0.97
54	76.1	79.8	0.95	86.5	88.9	0.97
55	75.1	78.7	0.95	85.5	91.3	0.93
56	75.7	86.3	0.98	83.8	89.6	0.93
57	72.5	76.3	0.95	82.3	90.6	0.91
58	70.1	77.1	0.91	79.5	86.4	0.92
59	66.4	72.2	0.91	78.6	85.6	0.91
60	64.0	85.0	0.86	76.0	84.8	0.89
61	60.5	69.2	0.87	72.2	81.5	0.88
62	57.2	72.0	0.79	69.5	79.1	0.87
63	53.3	63.0	0.84	66.5	77.5	0.85
64	49.6	57.8	0.85	63.2	74.1	0.85
65	46.0	57.7	0.79	56.8	69.3	0.81
66	41.3	49.9	0.82	53.2	66.6	0.79
67	37.8	56.4	0.67	51.7	63.0	0.81
68	33.4	48.3	0.69	46.5	59.8	0.77

* The table is continued in the next page. Entries are rates per 100 persons. *ad* is calculated as the ratio of survival rate in the population to the survival rate in the sample.

Table 2.2 (continued)

	Male			Female		
Age	Population	Sample	*ad*	Population	Sample	*ad*
69	29.3	46.3	0.63	42.5	61.0	0.69
70	25.8	49.1	0.52	35.1	50.4	0.69
71	22.2	35.4	0.62	30.8	46.0	0.66
72	19.3	36.1	0.53	26.2	44.8	0.58
73	15.2	28.2	0.54	25.2	47.2	0.53
74	13.7	26.5	0.51	21.3	42.1	0.51
75	10.1	35.1	0.28	16.2	38.3	0.42
76	7.2	24.1	0.29	14.0	42.5	0.32
77	3.1	12.5	0.24	10.1	32.7	0.31
78	4.3	18.5	0.23	8.7	30.0	0.29
79	3.4	17.1	0.19	7.1	30.8	0.23
80	2.8	38.9	0.07	5.8	21.9	0.26
81	1.1	5.9	0.18	4.9	28.3	0.17
82	1.7	17.4	0.10	4.2	27.6	0.15
83	1.1	5.1	0.21	3.6	20.0	0.18
84	1.5	10.0	0.15	1.5	8.3	0.18
85	1.4	20.0	0.07	2.1	21.4	0.10
86	1.1	8.8	0.12	1.1	6.3	0.17
87	1.1	8.8	0.12	1.3	11.1	0.11
88	1.1	8.2	0.13	1.2	33.3	0.04
89	1.1	33.3	0.03	1.2	22.2	0.05
90 or over	1.1	8.4	0.13	1.1	8.3	0.13

* Entries are percentages per 100 persons. *ad* is calculated as the ratio of survival rate in the population to the survival rate in the sample.

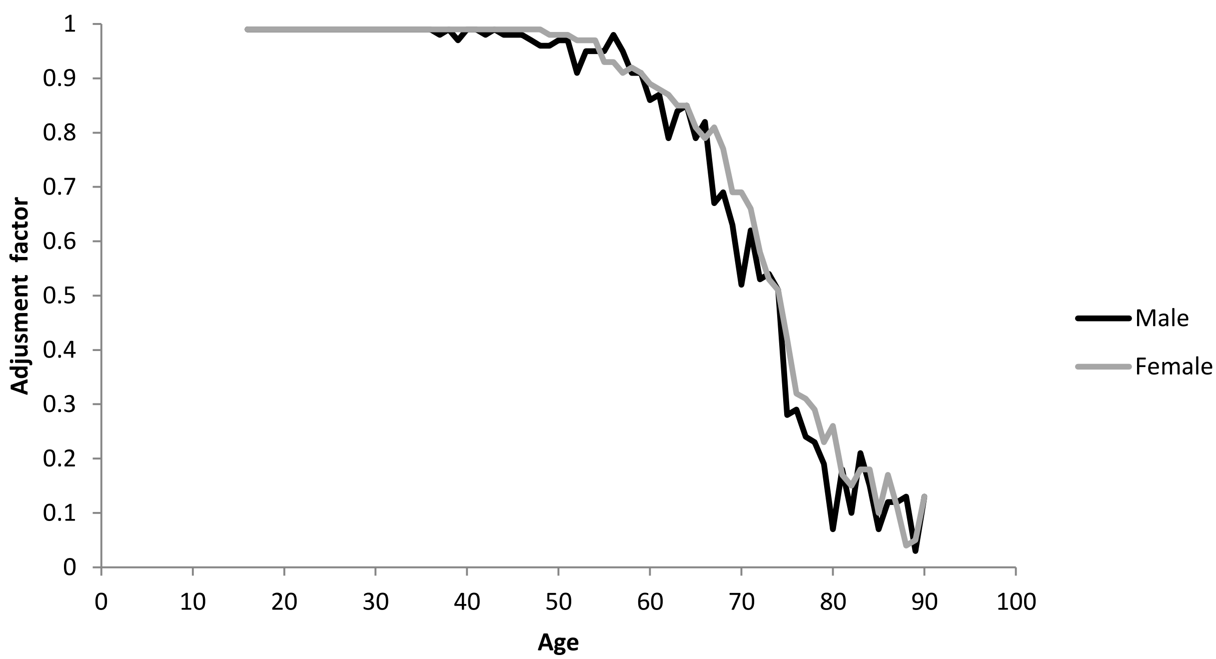

Figure 2.1: The adjustment factor by gender and age.

2.8.4 Weights creation

The standard weights (SWs)

The investigation in this chapter used waves 1 to 18 of the BHPS. To create the standard longitudinal weights at wave 18, we modelled the response at each wave conditional on responding in all of the previous waves. We started the modelling from wave 2 since the design weights in the BHPS – which will be multiplied by the *SWs*- are combined with wave 1 non-response weights. The model at each wave used variables from the previous wave. There were 17 models in total. Those who are known to be ineligible by wave 18 were not included in these models. Those whose eligibility is unknown by wave 18 were assumed as eligible cases and were included in the weighting models. In other words, we used the SWA illustrated by the model in equation 1.1 which was set out in chapter 1 of this thesis. For convenience, this model is rewritten as equation (2.10) below.

$$\mathit{Logit}\ \Pr(R_{i,t}=1/\ C_{i,t-1}=1)=f\ (\textstyle\sum_j \boldsymbol{\beta}_j Z_{ji}+\sum_k \boldsymbol{\beta}_k X_{ki,t-1}) \qquad (2.10)$$

Where t is the wave number for which the model is estimated (t=2, 3,…, T=18); i= 1, 2, …, $n_{1,..,t-1}$, where $n_{1,..,t-1}$ is the number of respondents who responded at every wave from 1 to *t-1* and who are known or assumed as eligible by the time of wave 18; $R_{i,t}$ is the response status at time (wave) t for respondent i ($R_{i,t}$=1 if response is observed at wave t; $R_{i,t}$=0 if response is not observed at wave t); $C_{i,t-1}$=1 if $R_{i,b}$=1 for all values of b from 1 to *t-1* (i.e. $C_{i,t-1}$=1 indicates that the model in wave t is conditioned on response in all of the previous waves); Z_{ji} is the set of time invariant variables for respondent i; $X_{ki,t-1}$ is the set of time variant variables for respondent i which are measured in wave *t-1*.

For the set of units who responded to all the 18 waves, the longitudinal *SWs* at wave 18 were calculated as the product of the inversed predicted probabilities from all 17 models, and wave 1 non-response/design weights (provided by BHPS) as shown in equation (2.11).

$$SW_i = D_i * \prod_{t=2}^{18} r_{ti}^{-1} \tag{2.11}$$

Where SW_i is the standard longitudinal weight at wave 18 for respondent i; r_{ti} is the predicted probability for respondent *i* from wave *t* model (*t*= 2, 3,…, 18);); i= 1,..., $n_{1,..,18}$ (where $n_{1,..,18}$ is the number of sample members who responded at every wave from 1 to 18); and D_i is wave 1 non-response/design weight for respondent *i*.

As covariates in each non-response model, we used the same variables that were used in the SWA in chapter 1 with the exception of health status since this variable will be used in our model of interest. Thus, the variables used are: gender, race, age, age-squared, tenure, presence of children in the household, education, type of household, employment status, type of house, number in full-time employment in household and region.

The final models are displayed in table (2.3). The table presents odds ratios. We interpret significance from $p<0.05$. Overall, the results here are in line with the SWA in chapter 1 and the general non-response literature. For example, most of the models here indicate that females, white sample members, homeowners and those who have more education are more likely to respond than males, non-white, non-homeowners and those with less education respectively. In contrast, most of the models suggest that sample members who are single-person household and those who live in an apartment building or other types of accommodations that are not a house are less likely to respond compared to sample members from multi-person households and those who live in houses respectively.

Table 2.3 Response propensity models based on the SWA (wave 2 to 18): modelling response in wave *t* conditional on responding in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9	Wave 10
Female	1.13*	1.39***	1.15	1.25*	1.29	1.16	1.55**	1.18	1.07
White	1.11	1.57**	1.71*	1.80*	1.21	1.42**	1.83*	1.13	1.67*
Age	1.05***	1.06***	1.07***	1.06***	1.12***	1.12***	1.13***	1.14***	1.13***
Age-squared	0.99***	0.99***	0.99***	0.99***	0.99***	0.99***	0.99***	0.99***	0.99***
Home owner	1.11	1.20**	1.46**	1.53***	1.28	1.32***	1.10	1.12	1.19
Has GCE qualification or above	0.97	1.33**	0.79	1.07	1.05	1.18	1.57***	1.27	1.44*
Employed	0.87*	0.92	1.01	1.15	1.16	1.21	1.45*	1.11	1.09
Others present in interview	0.95	0.78	1.05	0.99	1.66**	1.49**	1.17	1.08	0.62**
Single-person household	1.12	0.85**	0.73*	0.89	0.99	1.03	0.70*	1.24	0.62**
Household with children	1.50***	1.12	1.40*	1.31*	0.89	1.28	1.04	0.95	1.54*
Living in a flat	0.88	0.85	0.77	0.87	0.85	0.85	0.92	0.67*	0.70*
Living in other type of house	0.90	0.61*	0.60*	0.88	0.87	0.71	0.65	0.49**	0.92
1 or 2 persons in employment	0.79	.68***	0.81	0.96	0.94	1.23	0.94	1.05	0.88
3 + persons in employment	0.87	0.59*	0.77	0.85	1.01	0.66**	-1.06	1.09	1.05
South-East	1.11	0.80	1.57*	1.50*	1.01	1.25	0.73	1.13	1.36
South-West	1.16	1.01	1.17	1.17	1.84*	0.88	0.94	0.96	1.27
East Anglia	1.07	1.02	2.32*	2.08*	1.20	1.56	1.86	1.68	2.08*
The Midlands	1.06	0.70*	1.26	1.26	0.99	1.12	0.94	1.01	0.92
The North	1.28*	0.74	1.65**	1.19	0.89	1.71*	1.05	1.05	1.53
Wales	1.26*	0.70	1.84*	1.16	0.69	1.07	0.78	1.07	0.87
Scotland	1.06	0.59*	1.55*	1.51	0.38**	1.15	0.43**	1.06	1.26
N	8,706	7,906	7,483	7,172	6,850	6,571	6,282	6,132	5,931
Pseudo R^2	0.031	0.039	0.033	0.037	0.046	0.045	0.049	0.040	0.042

* The table is continued in the next page for the models of waves 11 to 18. The entries are odds ratios. In every wave response is modelled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. The reference categories of the categorical independent variables in the table are male, non-white, not a home owner, does not have a GCE or above degree, unemployed and others not present when interviewed, multi-person HH, household with no children, living in a house, no one is in employment in HH and London * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

Table 2.3 (continued)

	Wave 11	Wave 12	Wave 13	Wave 14	Wave 15	Wave 16	Wave 17	Wave 18
Female	1.05	1.26	1.50**	1.19	1.43*	1.04	1.36*	1.44*
White	1.87*	1.41	1.73*	1.61	1.69*	0.61	1.05	1.70
Age	1.12***	1.15***	1.14***	1.14***	1.21**	1.19*	1.16*	1.12
Age-squared	0.99***	0.99***	0.99***	0.99***	0.99**	0.99*	0.99*	0.99
Home owner	1.50**	1.42*	1.42*	1.45*	1.21	1.52*	1.19	1.33**
Has GCE qualification or above	1.12	1.60**	1.13	1.65**	1.15	1.20	1.18	1.16
Employed	1.15	1.13	1.65**	1.62*	1.22	0.98	0.61	1.39**
Others present in interview	0.90	0.93	1.11	0.86	0.97	0.98	1.31	0.91
Single-person household	0.67*	0.88	0.91	0.95	0.97	0.83	0.78	0.70*
Household with children	1.17	1.57***	1.13	1.62*	1.14	1.20	1.29	1.19
Living in a flat	1.03	0.62*	1.02	0.55**	0.79	1.43	0.97	0.94
Living in other type of house	0.92	0.67	0.58	1.37	0.78	0.96	0.95	1.27
1 or 2 persons in employment	0.80	0.80	1.11	0.93	1.55*	1.21	1.60*	0.58*
3 + persons in employment	0.32***	0.48**	1.13	1.21	1.92	1.41	1.38	0.43*
South-East	1.20	1.09	1.08	1.91*	1.88*	1.92*	1.54	1.74*
South-West	0.99	0.81	0.72	1.93*	1.98*	1.33	2.06*	1.50
East Anglia	1.81*	0.96	0.96	2.05*	2.63*	1.76*	2.15*	1.40
The Midlands	1.10	1.12	1.22	2.22**	1.30	1.62	1.31	2.01*
The North	1.28	0.99	0.77	1.69*	1.82*	2.13*	1.65	2.16***
Wales	0.94	1.32	0.80	2.17*	1.65	2.41*	0.96	2.58*
Scotland	1.01	0.75	1.26	1.47	1.10	1.35	1.53	1.23
N	5,781	5,605	5,456	5,340	5,223	4,654	4,554	4,310
Pseudo R^2	0.045	0.046	0.049	0.045	0.046	0.048	0.047	0.043

* The entries are odds ratios. In every wave response is modelled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. The reference categories of the categorical independent variables in the table are male, non-white, not a home owner, does not have a GCE or above degree, unemployed and others not present when interviewed, multi-person HH, household with no children, living in a house, no one is in employment in HH and London * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

For the responding cases, the *SWs* were then derived from the models displayed in table 2.3 (using equation 2.11). The distribution of these weights will be presented and discussed with the alternative weights in the next sections. However, as mentioned previously, there will be two sets of alternative weights. The first is AWs_1, and it comes from the main AWA introduced in this chapter (the *ad*-based method). The second is AWs_2, and it is based on the second alternative approach that was explained earlier (imputation-based method). Thus, we shall next describe how AWs_1 and AWs_2 were created before discussing the distributions of all the created sets of weights.

The first (main) set of alternative weights (AWs_1)

The set of AWs_1 was derived by adjusting the *SWs* using the values of the *ad* that were estimated previously (in table 2.2). This was done separately for men and women by a single-year age. For a given case *i*, which falls in the class of the gender *j* and age *k*, the AW_1 was calculated as the product of case *i SW* and the *ad* in the class of gender *j* and age *k*. This calculation is shown in equation (2.12).

$$AW_{1ijk} = SW_{ijk} * ad_{jk} \qquad (2.12)$$

Where *i* indicates the number of respondents (*i=1,2,…,4,097*); *j* is respondent's gender (*j=0, 1; where 0=male and 1=female*); *k* is the respondent's single year of age at wave 1 (*k=16, 17, …, 90+*); AW_{1ijk} is the first alternative weight of case *i* whose gender and age are *j* and *k* respectively; and SW_{ijk} is the standard weight of case *i* whose gender and age are *j* and *k* respectively.

Based on 2.12, the AW_1 of a given respondent is the modification of their *SW* according to the value of the *ad* in the class to which the respondent belongs. For example, the AW_1 of a female who responded in all of the 18 waves in our sample, and who was aged 16 at the time of wave 1 is $SW_{F,16}$* 0.99, where 0.99 is the *ad* for females aged 16 (from table 2.2). In contrast, the AW_1 of a female who was aged 85 at wave 1 is $W_{F,85}$* 0.10 (0.10 is the *ad* for females aged 85). Accordingly, as we discussed earlier, the weights' values for respondents in the panel who were aged under 60 at wave 1 will remain almost the same while the weights for those aged 60 or over change considerably. This comparison indicates that the AWA will have the same effect as the SWA on analyses restricted to respondents aged 16 to 59, whereas with analyses that are restricted to respondents aged 60 or over, or analyses on all respondents aged 16+, it may seem reasonable to expect a different impact on estimates if the AWA is implemented. In the latter case, the AWA will reduce the weights of respondents aged 60+, which in turn will change the distribution of the weights for the responding sample as a whole (likely to have smaller weight' values and variance), and may hence produce more precise and less biased estimates.

The second set of alternative weights (AWs_2)

The creation of the AWs_2 involved performing an imputation procedure to estimate eligibility status for those with unknown eligibility status by wave 18. Cases that were imputed as ineligible were excluded from the weights' construction. AWs_2 were then created using the same model-based method that we used to create the *SWs* with the exception of excluding the cases that were imputed as ineligible. To apply this approach we implemented the following:

By the time of wave 18, members of the original sample of the BHPS who responded in wave 1 are either known to be alive, known to be dead or of unknown eligibility status. This classification was set out when we calculated the sample survival rates in section 2.8.1. Based on this information, we created an indicator to reflect the eligibility status (ES) of every case of the original sample by the end of wave 18. For every case, ES takes one of three values: 1=eligible (alive), 0=ineligible[14] (dead) or .=eligibility is unknown. We then carried out random imputation on ES to impute the missing values (unknown eligibility statuses) as either 1 (eligible) or 0 (ineligible). The imputation was done using the *hotdeck* built-in command in Stata. This command performs random imputation which involves categorising the sample members into similar subgroups based on a number of specified variables. Missing values for sample members in any subgroup are randomly replaced with comparable values from sample members in the same subgroup. The variables used here were age (categorised)[15], gender, race, health status and financial situation from wave 1. These variables were used because they are available for all cases in the original sample, and because of their strong relationship with life expectancy (Rogers, Hummer and Nam, 2000; Singh-Manoux *et al,* 2008). Tables (2.4) and (2.5) present the results of this imputation.

By age at the start of the survey, table 2.4 shows the distribution of sample members with unknown eligibility status who were imputed as either eligible or ineligible cases. By wave 18 there are 3,118 cases of the original sample of the BHPS whose eligibility status is unknown. Recall that all of these cases were assumed as eligible cases and were used, along with those who are known to be eligible (5,588 respondents), for the weights'

[14] Note that the ineligible here include those who are institutionalised or emigrated.
[15] The age categories used were 16-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90+. It was necessary to use these age bands as they allow enough cases for imputation in every class constructed by the variables used in the imputation.

creation in the SWA. The imputation results here, however, suggests that 2,382 (76.4%) of those whose eligibility is unknown are eligible, whereas 736 (23.6%) cases are imputed as ineligible. Noticeably, in all ages under 60, more cases were imputed as eligible than ineligible. Conversely, in all age groups from 60+, more cases were imputed as ineligible than eligible. The majority of those who were imputed as ineligible are from age 60+ (453 cases). These results are remarkably consistent with our results from the adjustment factors suggesting that most of the ineligible sample members whose eligibility is unknown are clustered within one end of the age spectrum (aged 60+).

Table 2.5 presents the distribution of the eligibility status (ES) before and after the imputation. As can be seen from the table, the imputation increases the eligibility rate from 54.5% (observed) to 77.8% (observed + imputed). In turn, the 736 cases that were imputed as ineligible cases raised the ineligibility rate from 15.1% (observed) to 22.2% (observed + imputed).

Based on the imputation result, we created our second set of alternative weights (AWs_2) by restricting the modelling of the response propensity to 7,970 sample members who are either known or imputed as eligible cases. Those who were known, or imputed as ineligible cases, were excluded from the weighting (2,278 cases).

By excluding 23.6% (736 cases) of those whose eligibility is unknown, the values of AWs_2 are likely to be smaller in size compared to the SWs. This is especially in classes where more cases were imputed as ineligibles (those aged 60+ at wave 1). Since the weights of the AWA (AWs_1) are also expected to be smaller in size, for respondents who started the survey aged 60+, it might be reasonable to expect rather similar distributions of the AWs_1 and AWs_2. In contrast, the distributions of the AWs_1 and AWs_2 may be different than the distribution of the SWs which is likely to have larger weights especially for older respondents. In consequence, one may not expect larger differences between estimates constructed using AWs_1 and AWs_2. However, estimates constructed based on the SWs, where weights sizes may be mistakenly large for respondents aged 60+, might differ from estimates constructed based on AWs_1 and AWs_2. With the SWs, estimates are likely to be less precise and may be biased towards characteristics of older sample members (aged 60+).

Table 2.4 The distribution of the imputed cases whose eligibility is unknown by age at wave 1.

Age	**16-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-79**	**80-89**	**90+**	**Totals**
Eligible	872 (27.97%)	562 (18.02%)	479 (15.36%)	277 (8.88%)	143 (4.59%)	48 (1.54%)	1 (0.04%)	0 (0%)	**2,382** (76.4%)
Ineligible	41 (1.31%)	81 (2.60%)	74 (2.37%)	87 (2.79%)	174 (5.58%)	170 (5.45%)	98 (3.14%)	11 (0.36%)	**736** (23.6%)

* % were calculated out of the total number of the cases of unknown eligibility (3,118 cases).

Table 2.5 The effect of imputing eligibility status (ES): distribution of un-imputed and imputed ES.

	Un-imputed ES (observed)		Imputed ES	
	%	N	%	N
Eligible cases	54.5	5,588	77.8	7,970
Ineligibles cases	15.1	1,542	22.2	2,278
Unknown eligibility cases	30.4	3,118	-	-

*The imputation was done using *hotdeck* procedure. The variables used in the imputation were gender, age, race, health status and financial situation.

We applied the same models as the SWA (in equation 2.10) to model the response propensity for the purpose of deriving (AWs_2). The difference here is that we excluded the sample members who were imputed as ineligible cases.

The final models of the second alternative weighting are presented in table (2.6). Overall, the results are somewhat similar to those from the SWA, and also in line with the general literature of non-response. The distribution of the weights derived from these models (AWs_2) is discussed in the next section together with the distribution of the *SWs* and the AWs_1.

Table 2.6 Response propensity models of the second alternative approach: modelling response in wave *t* conditional on response in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9	Wave 10
Female	1.14*	1.29**	1.21	1.36**	1.02	1.41*	1.61**	0.92	0.93
White	1.71*	1.63**	1.60**	1.55***	1.07	1.74*	1.70	1.31	1.51*
Age	1.02**	1.04**	1.02*	1.3*	1.07*	1.01	1.01	1.08*	1.06
Age-squared	0.99**	0.99**	0.99*	0.96*	0.99*	1.00	0.99*	0.99	0.99
Home owner	1.08	1.32***	1.39*	1.54***	1.61**	1.29***	1.22	1.27	1.25
Has GCE qualification or above	1.37***	1.41**	1.13	1.17	1.01	1.47*	1.57***	1.43*	1.33**
Employed	0.87	1.07	1.15	1.11	1.09	1.54*	1.69*	1.12	1.43
Others present in interview	1.10	1.06	1.02	1.17	1.68***	1.55*	1.15	1.17	0.52***
Single-person household	0.96	1.25	0.58*	0.49*	0.89	0.63*	0.62*	1.20	1.09
Household with children	1.57***	1.06	1.34***	1.35*	1.13	1.57*	0.96	0.71	1.19**
Living in a flat	0.80*	0.79	0.67*	0.73*	1.10	1.07	0.79	0.45**	0.59*
Living in other type of house	0.82	0.46**	0.45**	1.22	0.95	0.92	0.79	0.37***	0.58*
1 or 2 persons in employment	1.12	0.74**	0.88	1.09	1.10	0.84	0.93	0.91	1.06
3 + persons in employment	0.97	1.16	1.08	0.89	0.64	1.19	0.59	0.80	0.61
South-East	0.91	0.84	1.74*	1.43	1.55	2.29**	0.34**	1.07	1.63
South-West	1.03	1.10	0.97	0.94	2.25*	1.08	0.45*	0.84	1.06
East Anglia	1.13	0.88	2.13*	1.58	2.14	2.93*	1.89	2.82*	3.22
The Midlands	0.94	0.76	1.06	1.15	0.97	1.57*	0.52*	1.18	1.10
The North	1.23**	0.84	1.63*	1.04	1.29	2.39**	0.83	1.11	2.15*
Wales	0.96	0.52**	2.84*	0.78	1.62	1.61	0.36*	1.03	0.61
Scotland	0.83	0.81	1.24	1.42	0.49*	1.711.	0.22***	1.33	1.28
N	7,970	7,270	6,946	6,770	6,606	6,435	6,265	6,090	5,913
Pseudo R^2	0.032	0.041	0.037	0.036	0.038	0.046	0.045	0.038	0.036

* The table is continued in the next page for the models of waves 11 to 18. The models do not include those who were imputed as ineligible cases. The entries are odds ratios. In every wave response is modelled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. The reference categories of the categorical independent variables in the table are male, non-white, not a home owner, does not have a GCE or above degree, unemployed and others not present when interviewed, multi-person HH, household with no children, living in a house, no one is in employment in HH and London * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

Table 2.6 (continued)

	Wave 11	Wave 12	Wave 13	Wave 14	Wave 15	Wave 16	Wave 17	Wave 18
Female	1.17	1.32	1.06	1.07	0.93	0.86	1.36	1.32
White	1.22**	0.74	1.87**	0.96	1.71**	0.22	1.22	1.26
Age	1.04	1.06	1.01	1.17***	1.14**	1.12*	1.02	0.97
Age-squared	0.99	0.99	0.98	0.99**	0.99**	0.99*	0.99	0.99
Home owner	1.29***	1.23	1.32**	1.57*	1.16	1.22	1.25	1.58*
Has GCE qualification or above	1.07	1.49*	1.09	1.46***	1.61**	1.15**	1.11	1.07
Employed	1.58**	0.95	0.93	0.98	0.65	0.80	0.67*	1.69
Others present in interview	1.04	0.90	1.01	1.07	1.03	1.82	1.16	0.96
Single-person household	0.57*	0.96	1.17	1.16	1.01	0.91	0.70	0.88
Household with children	0.85	1.12	1.26	1.03	0.95	0.69	1.17	0.86
Living in a flat	1.02	0.43**	1.13	0.75	1.14	1.31	0.79	1.17
Living in other type of house	0.93	0.54*	0.38**	1.12	0.97	0.63	2.49	1.81
1 or 2 persons in employment	0.51*	0.87	0.95	1.02	0.56*	1.01	1.12	0.54*
3 + persons in employment	0.25***	1.14	1.10	0.41*	0.36*	0.57	0.86	0.42*
South-East	1.97*	1.76*	1.46	4.33***	2.33**	3.46**	1.86*	1.94*
South-West	0.70	1.01	0.78	2.66**	2.53**	1.41	1.99	2.30**
East Anglia	1.79	1.72	0.64	6.44*	3.89*	6.14*	8.41*	1.65
The Midlands	1.30	1.64	1.07	3.47**	1.37	2.64**	1.29	3.16
The North	1.11	1.91*	0.93	193*	1.75*	2.91	1.55	2.46***
Wales	0.83	1.34	0.63	2.28*	1.87	6.21*	1.21	4.15
Scotland	0.98	1.27	1.53	2.31*	1.80	1.91	2.42*	1.10
N	5,773	5,583	5,440	5,285	5,197	4,596	4,375	4,189
Pseudo R^2	0.041	0.035	0.032	0.051	0.044	0.041	0.034	0.042

* The entries are odds ratios. In every wave response is modelled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. The reference categories of the categorical independent variables in the table are male, non-white, not a home owner, does not have a GCE or above degree, unemployed and others not present when interviewed, multi-person HH, household with no children, living in a house, no one is in employment in HH and London * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

2.9 Assessment of the AWA

Assessing the effect of the AWA requires comparisons between the AWs_1 and the weights resulting from the SWA (*SWs*). Aside from comparing the distributions of the weights themselves, the evaluation may involve constructing different types of estimates using *SWs* and AWs_1 and comparing them. During this process, the set of weights from the second alternative approach (AWs_2) can be used to check whether our introduced AWA is robust to changing the method used to estimate the eligibility of those with unknown eligibility status. Thus, our assessment of the AWA includes a discussion about the distribution of the *SWs*, AWs_1 and AWs_2. In addition, we used the three sets of weights in different types of weighted analyses. The latter includes producing descriptive statistics and estimating panel data models (multivariate analysis). The substantive analysis involved identifying the determinants of *Subjective Health Status* (SHS) since this is likely to be affected by the problem of including ineligible (deceased) cases in the SWA, which the AWA is designed to address.

We used a balance panel of those who responded in all of the 18 waves (4,097 respondents). Since the effect of the AWA is expected to be different for respondents who started the survey aged 60 or older than for those aged 16 to 59, the analysis was done separately for these two groups and for the full sample.

2.9.1 The distribution of the SWs, AWs_1 and AWs_2

We start the evaluation of the AWA by presenting the distribution of SWs, AWs_1 and AWs_2. Table (2.7) displays these, separately, for those who started the survey aged 16 to 59, 60+ and 16+ (all respondents in the panel).

For those aged 16-59, the three sets of weights have rather similar distributions, both in terms of central tendency and dispersion measures. This result is expected as both of our alternative weighting approaches are not expected to have weights that are considerably different than the standard weights for those aged 16-59. AWs_2, however, has less variability, indicated by the smaller Coefficient of Variation (CV), compared to AWs_1 and *SWs* (CV_{SWs}= 0.46; CV_{AWS1}= 0.46; and CV_{AWS2}= 0.37).

As for those aged 60+ and the full sample, the distributions of AWs_1 and AWs_2 remain fairly similar, but they differ from the distribution of the *SWs*. The weights mean value is bigger in the *SWs* ($Mean_{60+}$= 4.96; $Mean_{16+}$= 2.36) than in the AWs_1 ($Mean_{60+}$= 3.02; $Mean_{16+}$= 2.15) and in the AWs_2 ($Mean_{60+}$= 2.64; $Mean_{16+}$= 1.93). In addition, the dispersion is noticeably larger in the *SWs* than in the AWs_1 or the AWs_2[16] more so for those age 16+ than for those aged 60+. This is indicated by the larger values of the standard deviation (SD) and CV in the *SWs* (SD_{60+}= 4.52, SD_{16+}= 2.06; CV_{60+}= 0.91, CV_{16+}= 0.87) compared to the SD and CV in the AWs_1 (SD_{60+}= 2.03, SD_{16+}= 1.16; CV_{60+}= 0.67, CV_{16+}= 0.54) or the AWs_2 (SD_{60+}= 1.71, SD_{16+}= 0.88; CV_{60+}= 0.64, CV_{16+}= 0.46).

These results are in line with our expectation that the AWA (as based on the set of AWs_1) may have a different impact than the SWA, on estimates constructed from the set of respondents aged 60+ or estimates constructed based on all respondents aged 16+. Since the AWs_1 is both less affected by unknown eligibility and has less variability than the *SWs*, it is likely to produce more precise and less biased estimates. Additionally, since the

[16] Recall that there are fewer non-respondents in the weighting models of the second alternative approach as those imputed as ineligibles were excluded from these models which reduces the size of the weights for respondents with largest weights (those who are similar to non-respondents).

distribution of the AWs_1 is similar to the distribution of the AWs_2, across all sets of respondents, one may not expect the substantive results from the alternative weighting to be sensitive to changing the method of estimating the eligibility status for those whose eligibility is unknown.

Table 2.7 The distribution of SWs, AWs_1 and AWs_2.

	Respondents aged 16 to 59			Respondents aged 60+			All respondents aged 16+		
	SWs	AWs_1	AWs_2	SWs	AWs_1	AWs_2	SWs	AWs_1	AWs_2
Std.dev	0.96	0.94	0.68	4.52	2.03	1.71	2.06	1.16	0.88
Mean	2.07	2.03	1.85	4.96	3.02	2.64	2.36	2.15	1.93
CV	0.46	0.46	0.37	0.91	0.67	0.64	0.87	0.54	0.46
Min	0.41	0.41	0.35	1.24	1.06	0.85	0.41	0.40	0.35
Q1	1.49	1.47	1.40	2.49	1.99	1.52	1.53	1.50	1.33
Median	1.81	1.77	1.63	3.23	2.42	1.94	1.91	1.84	1.64
Q3	2.40	2.34	1.91	4.59	3.24	2.37	2.60	2.44	2.03
Max	10.94	10.83	8.89	49.18	25.08	19.89	49.18	25.08	19.89

*CV is the coefficient of variation (CV=Std.dev/Mean).

Turning to the substantive analysis, this was carried out to investigate subjective health status (SHS) in the BHPS using the SWs, AWs_1 and AWs_2. The following two sections summarise this analysis.

2.9.2 Descriptive statistics

In the BHPS, SHS is measured by asking respondents every year to rank their own health as excellent, good, fair, poor or very poor. The proportions of respondents in each of these categories are calculated using the SWs, AWs_1 and AWs_2 and are displayed in table 2.8. The weighted proportions are presented for our three sets of respondents separately (aged 16 to 59, 60+ and all respondents). To test whether there are differences between proportions constructed based on the alternative weights and the equivalent proportions constructed based on the standard weights, we used the built-in command *prtest* in Stata. *Prtest* performs a classical test of hypothesis on the equality of proportions. Using this command, the differences between the proportions calculated with the SWs were tested in turns with the equivalent proportions calculated with AWs_1 and AWs_2. The results of these tests are also included in table 2.8.

Focussing on the first set of respondents (aged 16 to 59) first, the SWs, AWs_1 and AWs_2 produced similar proportions across the categories of SHS, and all of the equality tests between the equivalent proportions do not show any significant differences. Thus, these results indicate that, for those who started the survey aged between 16 and 59, there are no significant differences between the proportions of SHS if the proportions are calculated using the SWs, AWs_1 or AWs_2. These results also confirm our expectation that the AWA does not change the standard weights of those who started the survey at a young age (between 16 and 59) because most ineligible sample members whose eligibility is unknown do not fall within this class of respondents. In addition, since AWs_2 also produced similar estimates as AWs_1, the results indicate that the AWA may be robust to changing the method that determines the eligibility status of those with unknown eligibility.

As for respondents who started the survey aged 60 or over, there are two significant differences here related to the categories 'poor health' and 'very poor health'. The proportions of these categories are significantly different (at the levels of *p< 0.01* and *p< 0.05* respectively) if calculated with *SWs*, than if they are calculated using the AWs_1 and AWs_2. The results show that these proportions are larger if calculated with *SWs*. These results clearly show the effect of including deceased sample members in the calculation of the *SWs*. Those deceased sample members, who were likely to be similar in their characteristics to older respondents with poor and very poor health, increased the sizes of the *SWs* of those respondents expanding their contribution to calculating the proportions of 'poor' and 'v. poor' health status. Furthermore, for the categories in question, the proportions calculated with the *SWs* have standard errors than the equivalent proportions calculated with the AWs_1 and AWs_2 indicating that the proportions calculated with AWs_1 or AWs_2 are more precise.

Finally, turning to all respondents in the panel, one significant difference is found. This is on the category 'v. poor'. The proportion of this category appears to be significantly different ($p< 0.05$) with the AWs_1 and AWs_2 than with the *SWs*. Furthermore, the proportion calculated with *SWs* has a larger standard error than the equivalent proportions calculated using the AWs_1 and AWs_2. Thus, the result here is in line, and has the same explanation as the previous results from those aged 60+. Also, both results confirm our expectation that the AWA may have a different effect on estimates than the SWA both on analysis restricted to older respondents and analysis based on the full sample. Moreover, the results from the AWA do not appear to be sensitive to changing the alternative weighting approach as AWs_1 and AWs_2 resulted in similar estimates.

Table 2.8 Weighted proportions across the categories of subjective health status using the SWs, AWs_1 and AWs_2.

SHS	Respondents aged 16 to 59			Respondents aged 60+			All respondents aged 16+		
	Using SWs	Using AWs_1	Using AWs_2	Using SWs	Using AWs_1	Using AWs_2	Using SWs	Using AWs_1	Using AWs_2
Excellent	23.16% *(.0071)*	23.21% *(.0070)*	23.38% *(.0070)*	16.56% *(.0166)*	18.74% *(.0174)*	18.07% *(.0171)*	21.48% *(.0064)*	22.18% *(.0065)*	23.05% *(.0066)*
Good	47.65% *(.0083)*	47.68% *(.0083)*	47.72% *(.0083)*	46.27% *(.0222)*	48.33% *(.0222)*	48.57% *(.0222)*	47.30% *(.0078)*	47.80% *(.0078)*	46.90% *(.0078)*
Fair	21.04% *(.0068)*	20.98% *(.0068)*	20.84% *(.0068)*	25.95% *(.0195)*	28.21% *(.0201)*	29.34% *(.0203)*	22.54% *(.0065)*	23.53% *(.0066)*	22.54% *(.0065)*
Poor	6.62% *(.0041)*	6.50% *(.0041)*	6.47% *(.0041)*	7.91%^{a} *(.0120)*	3.85%$^{a\,**}$ *(.0086)*	3.21%$^{a\,**}$ *(.0079)*	6.88% *(.0040)*	5.46% *(.0034)*	6.55% *(.0039)*
V. Poor	1.63% *(.0021)*	1.63% *(.0021)*	1.59% *(.0021)*	3.30%^{a} *(.0080)*	0.87%$^{a\,*}$ *(.00413)*	0.81%$^{a\,*}$ *(.0078)*	1.80%$^{a\,*}$ *(.0021)*	1.03%$^{a\,*}$ *(.0016)*	0.96%$^{a\,*}$ *(.0015)*

*The numbers in brackets are the standard errors. [a] indicates a significant difference between the proportions produced from the alternative weighting approaches and the corresponding proportion produced from the SWA. The differences between the proportions were tested using the command *prtest* in Stata. * $p< 0.05$ and ** $p< 0.01$.

2.9.3 Multivariate analysis

The multivariate analysis was carried out to investigate factors affecting SHS. This was done by estimating three groups of panel data models as will be explained next.

In this analysis, the five categories of SHS (excellent, good, fair, poor and very poor) were reorganised. The first three categories were combined into one category (good health status) and the last two were combined into another category (poor health status). Accordingly, SHS became a categorical variable with two categories, indicating whether the respondent has good or poor health status. This variable was used as the dependent variable in the analysis.

$$SHS_i = \begin{cases} 1, & if\ case\ i\ has\ a\ good\ health\ status. \\ 0, & if\ case\ i\ has\ a\ poor\ health\ status. \end{cases} \tag{2.13}$$

Where

SHS_i ≡Subjective health status.

The explanatory variables used here are gender, race, age, number of visits to the GP since last interview, smoking status, income, cohabitation status and financial situation. These variables are available across the 18 waves of the BHPS. Furthermore, they are known for their effect on health status and were used in prior research on self-assessed health in the BHPS (for example Jones *et al*, 2004).

We used a balanced panel from wave 1 to 18 (4,097 cases), and our three longitudinal sets of weights at wave 18 (SWs, AWs_1 and AWs_2) to estimate our substantive models. For each of our three sets of respondents (those who began the BHPS aged: 16 to 59, 60+ and 16+) we modelled SHS by estimating a random effects logistic regression model.

However, for each group, the model was estimated three times using the SWs, AWs_1 and AWs_2. This strategy allows the comparison between estimates resulting from each set of weights, separately for the three sets of respondents, while holding the estimation method constant.

To identify significant differences between equivalent coefficients estimated with the different weights, we conduct hypotheses testing on the differences between estimates adjusted with the *SWs* and their equivalent estimates adjusted with AWs_1 and AWs_2 using 95% Confidence Intervals (CI). Our test involves two essential steps. The first step is to construct 95% CIs of <u>the difference</u> between each two equivalent coefficients that are adjusted with the SWs and AWs_1 or AWs_2. Such CIs specify the range of values within which the difference between each two equivalent coefficients may lie. For example, if β_{SW}, β_{AW1} and β_{AW2} denote a given set of equivalent population parameters estimated by the equivalent set of coefficients b_{SW}, b_{AW1} and b_{AW2} which are adjusted with *SWs*, AWs_1 and AWs_2 respectively, we construct two CIs to test whether b_{SW} is different than b_{AW1} and/or b_{AW2}. These are CIs for (b_{SW} - b_{AW1}) and (b_{SW} - b_{AW2}). All CIs are 95% CIs, and are given by 2.14 below:

$$(b_{SW} - b_{AWi}) \pm 1.96 * S_{b_{SW}-b_{AWi}} \qquad (2.14)$$

Where $S_{b_{SW}-b_{AWi}}$ is the standard error of (b_{SW} - b_{AWi}) and is given by 3.9 below; and *i*=1,2.

$$S_{b_{SW}-b_{TWi}} = \sqrt{S^2(b_{SW}) + S^2(b_{AW}) - 2 * Cov(b_{SW}, b_{AWi})} \qquad (2.15)$$

Where $S^2(b_{SW})$ and $S^2(b_{AWi})$ are the variances of $b_{SW}\ and\ b_{AWi}$ respectively; $Cov(b_{SW}, b_{AWi})$ is the covariance of $b_{SW}\ and\ b_{AWi}$; and i=1,2.

The second step is to use the constructed CIs to test whether there is a significant difference between each two equivalent coefficients adjusted with the SWA and AWA (i.e. is there a significant difference between $b_{SW}\ and\ b_{AWi}$?). That is to test the following hypothesis:

H_0: β_{SW} - β_{AWi} =0 against H_a: β_{SW} - β_{AWi} ≠0; i=1,2.

Note that H_0 is rejected (i.e. there is a significant difference between $b_{SW}\ and\ b_{AWi}$) if the relevant CI does not include 0.

Table 2.9 presents the results of our substantive models. The table presents odds ratios and their standard errors. However, for ease of exposition, the 95% CIs that we used to test the differences between equivalent estimates adjusted with the different weights are not displayed in table 2.9. For all 95% CIs resulting from this analysis see appendix B.5.

Overall, although the models capture significant relationships between SHS and most of the factors included in the analysis, the importance of these factors differs between those aged between 16 and 59 and those aged 60+. For example, respondents from white ethnic groups are more likely to report good health status than respondents from non-white ethnic groups. However, this is not significant if the model is restricted to the set of respondents who started the survey aged 60+ ($\hat{b}_{60+,sw}$= 1.183, p> 0.05 and $\hat{b}_{60+,AW1}$= 1.692, p> 0.05; $\hat{b}_{60+,AW2}$= 1.855, p> 0.05). Nonetheless, our focus here is on the comparison between estimates resulting from the SWs, AWs_1 and AWs_2 within each group of respondents rather than differences in results across the different groups. Turning to this, the results of the comparison can be summarised in the following:

First, our attention is paid to the models of respondents aged between 16 and 59. Just as expected, the three sets of weights resulted in similar estimates here. The coefficients are similar both in terms of their magnitudes and standard errors resulting in identical significance levels across the three models.

Second, focussing on models for those aged 60+ and models for the full sample, some of the results generated from the *SWs* are different from the results derived from the AWs_1 and AWs_2. Overall, AWs_1 and AWs_2 produced rather similar estimates, suggesting that results from our introduced AWA might not be sensitive to changing the method used to estimate eligibility status for cases of unknown eligibility. Estimates resulting from the AWs_2 have, somewhat, smaller standard errors than the equivalent estimates resulting from AWs_1. However, this is expected since the dispersion in the distribution of the AWs_2 is lesser than in the distribution of the AWs_1.

On the other hand, comparing the estimates resulting from AWs_1 and AWs_2 with estimates resulting from the *SWs*, it can be noticed that, AWs_1 and AWs_2 produce more precise estimates compared with *SWs*. This difference is particularly clear for estimates based on those aged 60+ and less for estimates resulting from using all respondents aged 16+. With those aged 60+, the *AWs* reduced the standard errors substantially, for a number of variables resulting in an increase in their significance level. Namely these variables are: gender, age, number of visits to the GP since last year, smoking status, income and financial situation. Some of these variables, such as age, did not even appear to be significant with the *SWs*. Also, for estimates resulting from all respondents aged 16+, the AWs_1 and AWs_2 have effects on two variables: gender and income. Both variables are not significant with the *SWs*. With AWs_1 and AWs_2, however, the standard errors of these variables are reduced, which resulted in more statistical power for their corresponding coefficients suggesting that they are in fact significant at *0.05* level. Such variables are important in the process of understanding change in health status, and if their effects are interpreted incorrectly, different conclusions may be drawn. Thus, these results indicate the significant effect that the AWA may have on some of the survey estimates.

As for bias, this was checked by conducting our hypotheses testing (explained earlier) on the differences between equivalent estimates adjusted with the SWA and the AWT using the 95% CIs of the difference between the estimates (in appendix B.5). All tests indicate that the coefficients resulting from the SWA and AWA are not significantly different in terms of magnitude. This means that, unlike our descriptive statistics, our multivariate analysis does not show evidence of bias reduction.

To sum up, both the descriptive analysis and the multivariate analysis in this chapter indicate that the results from the SWA and AWA are similar in general. However, for certain types of estimates, the SWA may result in less precision and may produce biased estimates as a consequence of not dealing appropriately with the cases of unknown eligibility who are actually ineligible. This is especially likely in estimates related to the likely characteristics of ineligible sample members such as poorer health conditions. On the other hand, the AWA endeavours to eliminate the effect of including ineligible sample members in the estimation of the weights by implementing the adjustment procedure introduced in this chapter. The effect of the adjustment in the AWA is in a downwards direction reducing the incorrect sizes and the variance of the weights particularly for older respondents who started the survey aged 60 or over. As a result, estimates resulting from the AWA are more precise and, in some cases, less biased in comparison with the SWA. In addition, while the AWA produces - for some estimates - different results than the SWA, the results arrived at via the AWA do not seem to be sensitive to changing the weighting to the second alternative approach.

Table 2.9 Random effects logistic regression models for the determinants of good health status (subjective).

	Respondents aged 16 to 59			Respondents aged 60+			All respondents aged 16+		
	Using SWs	Using AWs_1	Using AWs_2	Using SWs	Using AWs_1	Using AWs_2	Using SWs	Using AWs_1	Using AWs_2
Year 1997 to 2002	0.941*(.038)*	0.938*(.035)*	0.938*(.035)*	0.993*(.091)*	1.037*(.070)*	1.078*(.061)*	0.950*(.035)*	0.961*(.032)*	0.982*(.030)*
Year 2003 to 2008	0.910*(.037)***	0.895*(.034)***	0.893*(.033)***	1.018*(.094)*	1.048*(.072)*	1.033*(.058)*	0.928*(.034)**	0.931*(.031)**	0.942*(.029)**
Female	1.110*(.073)*	1.128*(.070)*	1.128*(.070)*	0.558*(.100)***	0.621*(.083)****	0.507*(.055)****	0.991*(.062)*	0.981*(.055)**	0.893*(.047)**
White	1.685*(.207)****	1.767*(.201)****	1.790*(.203)****	1.183*(.673)*	1.692*(.345)*	1.855*(.327)*	1.677*(.205)****	1.729*(.194)****	1.728*(.191)****
Age	0.991*(.002)****	0.990*(.002)****	0.990*(.002)****	0.988*(.007)*	0.983*(.005)***	0.980*(.004)***	0.991*(.002)****	0.990*(.001)****	0.989*(.001)****
1 to 2 visits to GP since last year	0.352*(.033)****	0.357*(.031)****	0.357*(.031)****	0.639*(.112)**	0.609*(.078)****	0.511*(.056)****	0.392*(.032)****	0.408*(.029)****	0.396*(.027)****
3 to 5 visits to GP since last year	0.087*(.008)****	0.087*(.008)****	0.087*(.007)****	0.252*(.045)****	0.246*(.032)****	0.202*(.022)****	0.104*(.009)****	0.112*(.008)****	0.113*(.008)****
6 + visits to GP since last year	0.016*(.001)****	0.016*(.001)****	0.016*(.001)****	0.087*(.015)****	0.084*(.011)****	0.073*(.008)****	0.021*(.002)****	0.024*(.002)****	0.025*(.002)****
Smoker	0.696*(.039)****	0.699*(.037)****	0.700*(.037)****	0.639*(.110)***	0.647*(.083)****	0.635*(.072)****	0.686*(.037)****	0.688*(.033)****	0.683*(.032)****
Annual income/1000	1.007*(.002)****	1.007*(.002)****	1.007*(.002)****	0.979*(.006)***	0.976*(.004)****	0.975*(.004)****	1.002*(.003)*	1.003*(.002)**	1.005*(.002)**
Has a partner	0.983*(.123)*	0.989*(.099)*	0.925*(.077)*	1.016*(.054)**	1.019*(.051)**	1.020*(.050)**	1.027*(.050)*	1.042*(.045)*	1.039*(.042)*
Financially okay	0.860*(.045)***	0.872*(.043)***	0.874*(.042)***	0.727*(.082)***	0.707*(.061)****	0.626*(.045)****	0.840*(.040)****	0.838*(.035)****	0.795*(.032)****
Financially struggling	0.568*(.031)****	0.576*(.029)****	0.577*(.029)****	0.613*(.072)****	0.598*(.053)****	0.457*(.033)****	0.569*(.028)****	0.575*(.025)****	0.532*(.022)****
N	3,594	3,594	3,594	503	503	503	4,097	4,097	4,097
σ	1.60	1.60	1.61	1.69	1.72	1.72	1.62	1.62	1.62
ρ	0.44	0.44	0.44	0.46	0.47	0.47	0.44	0.45	0.45

* Entries are odds ratios. The numbers in brackets are the standard errors of the coefficients. The reference categories of the categorical independent variables are: year 1991 to 1996, male, non-white, no visits to the GP since last year, non-smoker, does not have a partner and having good financial situation. σ is the standard error of the random effects (sigma u). ρ is the percentage of the total variance that is due to differences between units.* p<0.05, ** p <0.01, ***p <0.001.

2.10 Conclusion

In this chapter we investigated a limitation in the SWA in relation to dealing with cases of unknown eligibility. The SWA assumes that cases whose eligibility is unknown are eligible, and therefore, it includes all these cases in the base of the model from which non-response weights are derived. This may be a rather ‘naïve’ method of handling unknown eligibility. It is unlikely that all cases whose eligibility is unknown are eligible, especially after many waves of data collection are conducted. Accordingly, if a large number of those whose eligibility is unknown are actually ineligible, weights resulting from the SWA may be incorrectly large in terms of their size and variance, and hence they may produce less precise and biased estimates.

The results from our investigation suggest the following:

Most of the original sample members of the BHPS whose eligibility is unknown by wave 18 are from the oldest age group in the sample, namely those who started the survey aged 60 or over. Accordingly, the adjustment made in the AWA affects the weights of those aged 60+ mostly, but as a result it changes the distribution of the weights for the sample as a whole. The weights resulting from the AWA are smaller in size (particularly the weights for older respondents) and have less variability compared to the standard weights. Thus, the resultant estimates from the AWA are more precise than the estimates produced from the SWA.

Despite differences in the weights distributions, overall, the SWA and AWA generate, rather, similar results. In general, for many of the estimates, the AWA do not change the conclusion arrived at via the SWA. However, for some estimates, the SWA and AWA yield different results. In these cases, the smaller standard errors associated with the AWA enable some estimates to be more significant in models that use alternative weights than in models that use standard weights. In a few cases, some estimates, which do not turn out to be significant with the SWA, appear to be significant with the AWA. It is clear that the SWA masks the significance of some estimates as a result of not handling unknown eligibility appropriately.

Moreover, supported by our descriptive statistics, the SWA may result in biased estimates. This is because most of the ineligible cases are clustered within certain classes in the sample (those who started the BHPS aged 60+). We found that the contribution of the larger standard weights of older sample members to some of the estimates was excessive. As a result, these estimates turned out to be significantly different from the equivalent estimates resulting from the AWA. With the SWA, the estimates in question contain influence of the ineligible sample members who are assumed as eligible, and who are similar in their characteristics to sample members who began the survey age 60+. Hence, these estimates were biased towards the characteristics of older sample members. Furthermore, with reference to the second alternative weighting, the results from the AWA do not seem to be greatly sensitive to changing the method of estimating eligibility status of sample members whose eligibility is unknown. Results from the AWA and the second alternative weighting are broadly similar.

The findings from this investigation have a number of important implications in the development of non-response weighting.

First, perhaps we downplay the overall role of the SWA. In principle, the SWA is a non-response weighting adjustment, which may be perfectly appropriate, particularly during the first few waves of the survey. Because, at the beginning of the survey, it might be plausible to assume that a large proportion of non-respondents are still eligible. Therefore, unknown eligibility may not be detrimental to weighting then.

However, as more waves are conducted, not only does the number of ineligible cases accumulate over time, it may also become a systematic feature of certain classes in the sample as it is shown by the distribution of our adjustments factor. It may still be possible to obtain valid conclusions by using the SWA in a number of analyses, but in certain types of investigations, the results might be misleading. For example, for analysts who would like to construct estimates of the longitudinal population at later waves, especially estimates related to older respondents, using the SWA may be at the cost of underestimating the importance of some of the factors in their analysis, or even obtaining biased estimates in some cases as was shown by our descriptive statistics here.

Second, and on the opposite side of the argument, the AWA offers a better system of dealing with uncertain eligibility compared to the SWA. It is convincing, robust and relatively easy to apply if information on eligibility is available at the population level. Furthermore, if the AWA is implemented, the interpretation of some weighted survey estimates may change considerably. In return, the new interpretations might change our understanding of the social process under investigation.

Third, although bias reduction was not established empirically by our multivariate analysis (only our descriptive statistics indicate this), we expect different results if the

AWA is applied on different data, where larger proportion of the unknown eligibility cases are ineligible. In this case evidence for bias reduction are likely to be clearer both on estimates from multivariate and descriptive analyses. Thus, for surveys that suffer from high rates of unknown eligibility, and where eligibility is also defined by being alive and living in the geographical area covered by the survey, the method is highly recommended.

However, when this approach is used, one should pay attention to the mortality rates that are used to calculate the adjustment factors. For accurate calculation of the adjustment factors, mortality rates should be up-to-date and reliable. For instance, the availability of mortality rates for the same population covered by the survey both in terms of time period and geographical area would improve the calculation of the adjustment factors. As an example, in this research, the BHPS sample was selected only from residential addresses, meanwhile registered population mortality statistics include people at all types of addresses (e.g. nursing homes). Thus, registered mortality rates may not perfectly match the rates in the population of interest, at least for the first two or three years of the survey (eventually, those initially institutionalised people will die, and all of the new institutionalised population will have been from the residential addresses covered by the survey, so at that point the survey should become representative of the entire population, i.e. the same population to which the mortality statistics refer). Moreover, the availability of population information on emigration and institutionalised individuals would be advantageous. Combining this information with mortality rates when calculating the adjustment factors will result in more accurate adjustment as all forms of ineligibility will be taken into account.

In surveys where ineligibility predominantly occurs by satisfying other characteristics (e.g. reaching a specific age or belonging to a certain social group) and maybe partially through death, the strategy of survival/death-based adjustment factor may not be very useful. This is because the calculation of the adjustment factor (which is based on comparing the survival proportions in the sample and the population) in this case will have not taken into account the main forms of ineligibility. The approach of the adjustment factor will be more effective if the main ineligibility form in the sample can be found in the records of population statistics, or other reliable external data, as the case in the current research (i.e. population mortality rates). Perhaps, in such a survey, the imputation-based approach (the second alternative) would be a better option.

Finally, another alternative procedure (used in the Health Survey for England and HILDA) could be contacting the death register office. In almost every country there is an office where deaths are registered. These offices collect information such as name, time and date of death, place and date of birth, the last address, occupation, reason for death and contact information of a surviving person related to the person died (usually a spouse or civil partner). If the survey organisation is able to contact the death register office and obtain this information, death can be identified by matching the records of respondents of unknown eligibility with the information held in the register office. The advantage of this approach is that it produces precise estimates based on accurate information. However, apart from the fact that this approach is time consuming (need to be done for all unknown eligibility cases at every wave), in some countries, register offices may not be willing to co-operate, for reasons of confidentiality.

In any case, for panel household surveys that are similar to the BHPS, any method for dealing with unknown eligibility in the context of weighting should focus on sample members who started the survey at older ages (may be aged 60+) as most of the ineligible (deceased) cases are likely to be centred within this age group.

CHAPTER 3

Non-response Subgroup-tailored Weighting: the Choice of Variables and the Set of Respondents Used to Estimate the Weighting Model

3.1 Introduction

In panel studies, the use of the logistic regression model to predict the probability of response and create non-response weights is classic. In most cases, the model is estimated using typical weighting variables (such as age, gender, race, etc…) and all cases in the selected sample for which data is available on the weighting variables. This is a typical feature of the SWA in longitudinal surveys which was described in the main introduction to this thesis. Since all sample members are used in the process of modelling the response propensity and deriving the weights, it may be necessary for the SWA to use 'generic' weighting variables. These are variables that are successful in predicting response for the sample as a whole and, also, may be correlated with some of the survey key variables. Consequently, variables that only distinguish response from non-response at a sub-group level may not be used in the SWA if they do not appear important at the full sample level.

It is important to use variables that are correlated with the survey target variables in order to produce a set of weights that is successful in reducing non-response bias (Särndal and Lundstrom, 2005; Little and Vartivarian, 2003, Kreuter and Olson, 2011). However, the extent to which the bias is reduced is also based on a good specification of the model in terms of using variables that significantly explain the variation in the response propensity in all sub-groups in the sample. Otherwise, the weights will not reduce non-response bias in estimates related to sub-groups where variation in response propensity either is not, or is poorly, accounted for by the weighting variables. Moreover, the weights will reduce non-response bias to the maximum possible extent if they are used to adjust an estimate that is constructed using the set of respondents used to create the weights.

In practice, it is unlikely for the SWA to be able to account for the variation in the response propensity in all sub-groups in the sample, given that it is based on just one weighting model, using all sample members and common weighting variables. Because, even in the same survey sample, the phenomena that cause non-response can differ across different sub-groups, both in terms of scale and type.

For example, consider a survey that collects data from individuals belonging to different social classes. For a particular social class, say one that is formed of teachers and lecturers, assume that there is a rate of non-response amongst this group. It may be plausible to assume that non-response rate in the sub-group in question is low compared to non-response rates in other sub-groups belonging to other social classes in the sample. This is because, individuals within academia may feel obligated to cooperate with the survey out of their academic scene of duty. In any case, in this example, it is likely that the factors responsible for non-response in the sub-group of teachers and lecturers are rather different than the usual non-response predictors (such as age, gender, race and education), which could be more responsible for non-response in other sub-groups in the sample. Meanwhile, with teachers and lecturers, variables such as age, gender, race and education might not explain much of the variation in the response propensity.

In the light of this scenario, consider a case in which a researcher would like to construct an estimate using only the set of teachers and lecturers in the sample. However, the researcher decides to use non-response weights to reduce any potential bias in the estimate in question. Thus, a model that is correctly specified to predict response probability in general (i.e. SWA) which is based upon all sample units, using variables that may be strongly correlated with the response propensity in many sub-groups in the sample but weakly correlated with the response propensity in the sub-group of teachers and lecturers, might result in a set of weights that successfully reduces non-response bias in many survey estimates but not necessarily in estimates which are constructed using the set of teachers and lecturers. With respect to any analysis that is restricted to this sub-group, weights would be more effective if the weighting model is:

a) Estimated to deliberately account for the variance in the response probability in the sub-group of teachers and lectures by using variables that strongly affect their response propensity regardless of whether or not they also affect the response propensity in other sub-groups in the sample.
b) Estimated using the set of teachers and lecturers only.

Since it is unlikely for a SWA to predict response in all sub-groups in the sample, an alternative weighting approach could set a modelling strategy that is able to account for the variation in the response propensity in a selected set of sub-groups. In this approach a number of different weighting models can be estimated with an intention to: explain a larger proportion of variance in response propensity in certain sub-groups in the sample (perhaps some of the main sub-groups in the sample which are more likely to be used for analysis); use a particular set of variables (rather than generic) which account for variation in the response probability in these sub-groups; and estimate the model by using sample members from the sub-groups in question only (i.e. by using the set of respondents that analysts will likely to use to construct estimates of interest). This way the weights derived from each weighting model can be more powerful in dealing with non-response bias in their relevant sub-groups in comparison with weights derived from the SWA. In addition, if the sub-groups selected for this type of weighting represent some of the major domains in the sample, the resultant weights may also reduce non-response bias in estimates constructed based on the whole sample (total sample estimates) if they are put together appropriately. This strategy of weighting is discussed as the AWA in this chapter.

The chapter investigates whether there is evidence to show that designing weights for specific sub-groups in the sample can significantly affect survey-based estimates from these sub-groups to the extent that they become different from the estimates produced through the SWA. The introduced AWA will be referred to as 'subgroup-tailored weighting approach' (S-TWA) and weights produced from this approach will be called 'tailored weights' (*TWs*).

The BHPS sample will be used to study the differences between the SWA and the proposed S-TWA. As in the previous chapters, the investigation here is based on creating weights using the SWA (*SWs*) and weights based on the AWA (*TWs*), and compare estimates resulting from a substantive analysis that uses the *SWs* and *TWs.*

The idea of the S-TWA will be investigated by selecting two sub-groups, from the BHPS sample, on which substantive analyses are intended to be done. The tailored weights will then be designed for these sub-groups by using variables that are thought to be associated with the response propensity in the sub-groups under investigation regardless of whether or not these variables are also used in the SWA. This means that the S-TWA will add new variables to the common variables that are usually used in the SWA to create the *TWs.* The new variables will be considered under the assumption that they are important predictors of the response in the selected sub-groups even if they are not important in terms of predicting response in other sub-groups. In turn, some of the variables used in the SWA may not be used in the S-TWA if they do not distinguish response from non-response in the sub-groups in question even if they are important in terms of predicting response in other sub-groups. Additionally, the tailored weights will also be created by restricting the weighting models to the sets of sample members in the selected sub-groups.

3.2 The choice of sub-groups

The data used in this chapter were from the first eight waves of the BHPS[17]. The data cover the period 1991 to 1998. The analysis was restricted to sample members who responded at wave 1, and who were aged 16 or older at that time. The tailored weights (*TWs*) were designed to deliberately target non-response bias in estimates related to two sub-groups of sample members:

1) Those who retired in the year 1991 or before.
2) Those who were born in the year 1965 or after.

The first sub-group refers to the group of retired sample members (relatively old sample members). These are sample members who started the survey as retired individuals. Therefore, this sub-group does not include respondents who retired at a later wave (i.e. in any year from 1992 to 1998). Since the analysis is restricted to sample members aged 16 or older, the second sub-group identifies sample members who were within the age group 16 to 26 at the start of the survey (younger sample members).

[17] Some of the variables used in the analysis are not available across all waves.

While other types of sub-groups in the BHPS sample are important too (e.g. those who were born outside the UK or the set of disabled sample members), the selected sub-groups here represent major domains in the sample. Also, both of these sub-groups contain enough sample members to allow valid investigation of the issue discussed in this chapter. In addition, both sub-groups, together, include a balanced set of sample members in terms of gender, age (young and relatively old respondents) and labour market status (out of the labour force and working age individuals). Furthermore, a large number of substantive analyses may be conducted on the selected sub-groups. Thus, whether a set of weights that is tailored to these sub-groups results in different estimates than estimates produced with the SWA is worth investigating.

The sub-group choice in our analysis splits the sample into three non-overlapping sub-groups: 1) retired sample members; 2) sample members who were born in 1965 or after; and 3) non-retired sample members who were born before 1965 (i.e. the rest of the sample). However, the S-TWA focuses on retired respondents and those who were born in 1965 or after. The tailored weights will be created to adjust for the longitudinal non-response up to wave 8. Thus, the weights will be appropriate for analyses, on the selected sub-groups, that use a balanced panel from the first eight waves of the BHPS. Additionally, since we are using two of the major sub-groups in the sample, the weights are also likely to reduce non-response error in estimates related to full sample analyses.

Before describing how the tailored weighting is done, the next section will briefly outlines the construction of the *SWs.*

3.3 Weights from the SWA (*SWs*)

Constructing the *SWs* for this investigation was a typical SWA. In this chapter, we apply the same SWA that we applied in chapter 1 and 2 which was done on the basis of the SWA that we set out in chapter 1 (given by equation 1.1). The difference here is that we only model the response propensity up to wave 8, whereas in chapter 1 and 2 it was done up to wave 15 and 18 respectively. For convenience, we re-explain this process in what follows:

There are eight waves in total for the current analysis (from wave 1 to 8). The *SWs* were created to adjust for the longitudinal non-response at wave 8. The process involved modelling the response propensity at each wave conditional on responding at all of the previous waves. Those who are known to be ineligible by wave 8 were not included in these models. Those whose eligibility is unknown by wave 8 were assumed as eligible cases and were included in the weighting models. Consequently, the analysis was restricted to sample members known (or assumed) to be part of the target longitudinal population at wave 8 which the weighting here aims to represent. The model at each wave used variables from the previous wave. The variables used to model the response propensity are the usual weighting variables in the SWA. Namely these are: gender, race, age, age-squared, tenure, presence of children in the household, education, type of household, employment status, type of house, number in full-time employment in household and region.

We started modelling the response propensity from wave 2 as the BHPS offers wave 1 non-response weights combined with the design weights. Equation (3.1) below explains this process in notations.

$$\textit{Logit}\ \Pr(R_{i,t}=1/\ C_{i,t-1}=1)=f\ (\sum_j \boldsymbol{\beta}_j Z_{ji}+\sum_k \boldsymbol{\beta}_k X_{ki,t-1}) \qquad (3.1)$$

Where t is the wave number for which the model is estimated (t=2, 3,…, T=8); i= 1, 2, …, $n_{1,..,t-1}$, where $n_{1,..,t-1}$ is the number of respondents who responded at every wave from 1 to *t-1* and who are known or assumed as eligible by the time of wave 8; $R_{i,t}$ is the response status at time (wave) t for respondent i ($R_{i,t}$=1 if response is observed at wave t; $R_{i,t}$=0 if response is not observed at wave t); $C_{i,t-1}$=1 if $R_{i,b}$=1 for all values of b from 1 to *t-1* (i.e. $C_{i,t-1}$=1 indicates that the model in wave t is conditioned on response in all of the previous waves); Z_{ji} is the set of time invariant variables for respondent i; $X_{ki,t-1}$ is the set of time variant variables for respondent i which are measured in wave *t-1*.

Table (3.1) displays the results of the final models of the SWA.

Table 3.1 Response propensity models based on the SWA (wave 2 to 8): modelling response in wave *t* conditional on responding in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Female	1.31*	1.22**	1.03	1.36*	1.13	1.29**	1.41**
White	1.69*	1.80**	1.23	1.71*	1.05	1.66*	1.83**
Age	1.07***	1.08***	1.10**	1.08**	1.11**	1.09**	1.12***
Age-squared	0.99***	0.99***	0.99**	0.99***	0.99**	0.99**	0.99**
Home owner	1.43*	1.39**	1.28*	1.08	1.03	1.04	1.58**
Has GCE qualification or above	1.21*	1.04	1.13*	1.10	0.92	1.26*	1.19**
Employed	0.89*	0.79	1.16*	1.24*	0.87	1.39	1.31
Others present in interview	0.87	1.28*	1.07	1.22	1.37*	1.19	0.94
Single-person household	0.72	1.18	0.76*	0.94	0.88*	0.91	1.08
Household with children	1.39*	1.44*	0.89	1.06	1.58*	1.08	1.43
Living in a flat	0.98	0.90	0.90	0.88	0.91*	0.87*	1.10
Living in other type of house	1.12	0.91	0.75**	0.93	0.64*	1.21	0.87
1 or 2 persons in employment	1.11	0.79*	0.94	0.89	1.22	1.31	1.07
3 + persons in employment	0.93	0.63*	0.89	0.61*	1.12	1.16	0.88*
South-East	0.93	1.27	1.86*	0.84	1.33	1.45*	1.42
South-West	0.96	0.95	1.24	1.44	0.92	1.08	1.25
East Anglia	1.03	0.88	2.03*	1.86*	0.96	1.14	1.28
The Midlands	0.86	1.59	1.76	1.07	0.89	0.93	1.10
The North	1.23**	0.72*	1.51*	1.22	0.87	1.48*	1.36
Wales	1.44*	0.88	1.30	0.84	0.63*	1.26	0.79
Scotland	0.91	1,33	0.85	1.72*	0.49*	1.24	0.61*
N	9,593	8,699	8,218	7,863	7,496	7,152	6,878
Pseudo R^2	0.031	0.032	0.038	0.035	0.034	0.33	0.035

*The entries are odds ratios. In every wave response is modelled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. The reference categories of the categorical independent variables in the table are male, non-white, not a home owner, does not have a GCE or above degree, unemployed, others not present when interviewed, multi-person HH, household with no children, living in a house, no one is in employment in HH and London * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

For the set of responding sample members in the 8 waves, the longitudinal *SWs* at wave 8 were calculated as the product of the inversed predicted probabilities from the models in table (3.2), and wave 1 non-response/design weights (provided by BHPS) as shown in equation (3.2).

$$SW_i = D_i * \prod_{t=2}^{8} r_{ti}^{-1} \tag{3.2}$$

Where SW_i is the standard longitudinal weight at wave 8 for respondent i; r_{ti} is the predicted probability for respondent *i* from wave *t* model (*t*= 2, 3,…, 8);); i= 1,..., $n_{1,..,8}$ (where $n_{1,..,8}$ is the number of sample members who responded at every wave from 1 to 8); and D_i is wave 1 non-response/design weight for respondent *i.*

The distribution of the *SWs* is presented and discussed later on with the *TWs.*

3.4 Proposed weighting variables for the subgroup-tailored weighting

Aside from the variables that are used in the SWA, for each of our selected sub-groups, some variables may be of a particular interest in terms of predicting response in the sub-groups under investigation. These variables are not used in the SWA as they do not associate with the response propensity for the sample as a whole. In this section, we shall propose and discuss two sets of these variables that will be used to create the tailored weights for our two sub-groups. In the next section we explain the methodology that will be implemented in the S-TWA.

3.4.1 Proposed weighting variables for r*etired sample members*

1- Religion: having a religion is considered as a form of social participation. While some research suggests that social participation can negatively affect the contact attempt –by affecting the at-home pattern- (Lepkowski and Couper, 2002), other research supports the idea that social participation is an indication of higher human interaction levels and therefore a person who is socially interactive is more likely to cooperate and provide data for the survey (Groves and Couper, 1998). As for the BHPS sample, Uhrig (2008) found that those who have religious beliefs are significantly more likely to respond than those who do not have religious beliefs. However, he found that this significant effect disappeared once other variables such as organisational participation (e.g. joining sport clubs and professional organisations) are included in the model. This is because organisational participation is also an indicator of higher human interaction levels and hence survey cooperation. However, some of the organisational participations are more common among working-age respondents than retirees especially if they require a high load of physical activities and/or someone within the labour force. In this research, it was assumed that organisational participation such as joining sport clubs and professional organisations is more common amongst working-age respondents than their retired counterparts and hence it can only affect the estimated association between religion beliefs and survey participation of working-age respondents. As for the retired respondents, religion can then be considered as a good predictor of non-response.

2- Respondent's energy compared to average at their age: the effect of this variable on response propensity can be viewed in two different ways. On the one hand, those who are more energetic than average at their age can be more mobile and are less likely to stay at home than those who have less energy. Thus, for surveys that make contact with respondents at their homes, it is more likely to find less energetic people at home than those with more energy. On the other hand, having less energy than average at their age may be associated with bad health conditions implying a lower level of cooperation or even refusal due to health conditions. Prior research on non-response suggests that refusal for health reasons is common amongst elderly respondents (Uhrig, 2008). For the sub-group of retired respondents (relatively old sample members), energy compared to average at the same age can be seen as an important indicator for both at-home pattern and health condition. Thus, whether or not this variable affects response propensity in the sub-group of retired respondents is worth investigation.

3- Whether respondent supports a political party: there is little research that has used political views and opinions to predict non-response since it is not clear that there is a direct relationship between the two factors. However, some of the literature in this area (e.g. Groves and Couper, 1998) implicitly indicate that those who have political views, such as supporting a political party, may be more aware of the government's role in the society and therefore may feel more obligated to provide data for the survey. Some of the literature on political engagement suggests that it is lower amongst working-age persons. One reason for this is that working-age respondents often do not have the time to engage with politics (Brandon, 2012). On the other hand, retirees do not often face time problems; instead, they have the time to participate in politics. In fact, retirees may feel the need to be socially interactive and therefore may participate in politics. Moreover, retirees could support and vote for a political party for reasons such as protecting the valuable benefits they receive from the government. Thus, based on the assumption that supporting a political party can influence response propensity and it is more frequent amongst retired respondents (Brandon, 2012), this variable was considered as a good weighting variable for retired respondents.

4- Subjective financial situation: research on non-response has established the positive relationship between wealth/financial position and response propensity (Groves and Couper, 1998; Fitzgerald *et al.*, 1998; Lepkowski and Couper, 2002). That is to say, those who are in better financial positions are more likely to respond than those who are less well off. However, for the BHPS sample, the evidence for subjective financial situation is in contradiction with the general financial findings. Previous research on subjective financial position on the BHPS has found that those who subjectively report themselves as being in better financial positions are less likely to respond than those who report themselves as being in worse financial positions (Uhrig, 2008). Nonetheless, the effect of subjective financial situation might change and confirm evidence from the general non-response literature once some sub-groups in the sample are controlled for (i.e. when the investigation is only done on retirees for example). For this reason, subjective financial situation was added to the set of weighting variables of retired respondents.

5- Having access to a car: having access to a car for personal use is considered –to an extent- as an indication of wealth and a good financial situation (Uhrig, 2008). As for retired respondents, having access to a car may also be thought of – to some extent - as an indicator of a good health (since driving a car require performing a set of physical acts that may not be possible to conduct with a bad health condition). Thus, this variable was added here under the assumption that it is indicative of good health status and good financial situation.

3.4.2 Proposed weighting variables for *those who were born in 1965 or after*

1- Liking the neighbourhood: this, in a way, expresses whether one is attached to one's current neighbourhood. The feelings of respondents about their settlement in a neighbourhood are indicative of whether they will continue to live in that neighbourhood, and hence of the likelihood of locating and contacting them successfully. There is evidence in the literature that younger respondents are more likely to move house (Uhrig, 2008). Thus, this variable is likely to have a distinctive effect on the response propensity for those who were born in 1965 or after (younger respondents) compared to their counterparts' sub-groups. Thus, this variable was added to the weighting variables of this sub-group.

2- School leaving age: it is well known that in the United Kingdom (UK) most people leave school at the age of 15 or 16. However, there are some exceptions where people may leave school at different ages, either aged less than 15 or more than 16. This may occur, for example, due to coming to the UK at the age of six and having to start school a year or two later than the average starting age (five years old). Circumstances in which one has to leave school at a different age than the average person may affect one's tendency to participate in the survey. Regardless of the nature of these circumstances, their existence can be expressed through the school leaving age. In this analysis, it is assumed that the effect of the circumstances associated with the school leaving age on survey participation fades over time. In other words, the effect is stronger at a younger age than at an older age. This is because living longer enables one to experience more life-events that may reduce any influence on survey cooperation due to the reasons why they left school at a different age than the average person. Thus, the relationship between school leaving age and non-response maybe of more interest for those who were born in 1965 or after than for those who were born before 1965.

3- Having children: this measures whether the respondent has his or her own children within the household. Non-response theory suggests that the presence of children in the household is positively associated with survey response (Groves and Couper, 1998; Lepkowski and Couper, 2002; Uhrig, 2008). This is regardless of whether or not these children are the respondent's own children. Because, households with children are more settled and less likely to move house, and even if they move house, they are easier to relocate and contact since there are children in the household. This is especially important for younger respondents who are more mobile and less settled. Therefore, an item that measures if the respondent has their own children within the household for those who were born in 1965 or after (younger respondents) can be considered as a good weighting variable for this sub-group. This is because of its distinctive effect on the response process of those who were born in 1965 or after.

4- Subjective financial situation: it was mentioned earlier that the evidence for subjective financial situation in the BHPS is in contradiction with the general financial findings (in the BHPS those in better financial positions are less likely to respond than those in worse financial positions). Thus, similar to the sub-group of retired respondents, it is worth testing the effect of subjective financial situation on the response propensity of those who were born in 1965 or after too.

5- Having access to car: aside from being indicative of wealth, having access to a car may have a distinctive effect on younger survey participants. It can be argued that having access to a car may affect the contactability of younger respondents. Therefore, this was included in the set of weighting variables of those who were born in 1965 or after.

3.5 The subgroup-tailored weighting approach (S-TWA)

In this analysis, our aim is to incorporate the proposed weighting variables to construct a set of weights that is tailored to two sub-groups in the sample: retired sample members and sample members who were born in 1965 or after. There are at least two ways to do this:

Interaction-based approach: with this approach the response propensity can be modelled as done in the SWA, but interactions of the proposed variables for tailored weighting for the two sub-groups under investigation will be added to the models. For example, to capture the effect of religion (one of the proposed variable for the S-TWA) on the response propensity of retired sample members, one may add an interaction term of the variable that indicates whether a sample member is retired, and the variable that measures religion, to all of the weighting models estimated in the SWA.

Testing interaction effects may be a task that should be performed when a response propensity model is estimated. However, including interactions in the response propensity models that are used to derive non-response weights is not a common practice amongst survey researchers (Brick, 2013)[18]. Moreover, it is certainly not one of the features of the SWA that we set out in the introduction of our thesis. Most survey organisations tend to rely on main effects when estimating their response propensity models for construction of weights. Furthermore, even in cases where interactions were

[18] Unlike panel studies, some cohort studies such as the 1958 National Child Development Study (NCDS) in the UK used interactions to model the response propensities (Hawkes and Plewis, 2006).

used, the results suggest that weighting models with interaction effects have similar outcomes to weighting models with only main effects (e.g. Schouten, 2004). Perhaps this is because, even when the interaction effects are used, they are only considered between the standard non-response variables (variables that affect response probability for most sample members), rather than considering interactions between standard variables and variables that only predict non-response at a sub-group level.

For example, as explained in the introduction of the thesis, the weighting in the BHPS was done, at every wave, using a weighting class method. In each class, the responding cases were weighted by a factor that made their total number equal to the responding and non-responding cases in the class. An automatic interaction detection programme (CHID) was used to define the weighting classes[19]. These classes are equivalent to interaction terms included in a response propensity model. However, the method relied on a number of variables that were thought to be informative of non-response in the BHPS sample as a whole and of critical interest in the analysis of the BHPS data (i.e. standard variables). At every wave, CHID was used to detect important interactions (in terms of response) amongst the predictor variables and define the classes. For instance, at wave 18, the candidate predictor variables used for weighting were: gender, age, race, tenure, health status, employment status, type of household, type of accommodation, region, household size, education, income, number of rooms in the accommodation and whether there is a dish washer in the accommodation. Important interactions in terms of predicting response were: age*gender, age*region, race*employment status,

[19] CHAID is a statistical tool used for segmenting a population in terms of some dependent variable (in our cases the probability of response) using a set of predictor variables. Predictor variables are typically categorical. It successively scans all the variables in the predictor set to identify the ones which best discriminate cases on the basis of values of the dependent variable. It uses these variables to categorise the cases in the sample into a number of classes based on a user defined minimum class size.

income*region, age*race, tenure*whether there is a dish washer in the accommodation, household size*number of rooms in the accommodation, type of accommodation*region, health status*education and household size*employment status.

For our subgroup-tailored weighting, after including all the necessary interactions in the weighting models, the tailored weights can then be calculated, as usual, as the product of the inversed predicted probabilities from the estimated models. The advantage of this approach is that it is relatively straightforward to apply. Also, it results in a single set of weights that is tailored to retired respondents and those who were born in 1965 or after. However, it may have some disadvantages. First, if there are many variables suggested for the S-TWA for each sub-group, the number of interaction terms becomes impractically large to include in one model. This is especially if some of the proposed variables for the S-TWA are categorical variables with many categories (more than 2 categories). If too many interactions are included in the weighting model, this may result in less statistical power for other important variables in the model. Second, it uses all sample members to model the response propensity, including those who are not in the sub-groups under investigation. Thus, some variables, from the SWA, which are not associated with the response propensity in the sub-groups in question, will be kept in the weighting models because they may be correlated with the response probability in other sub-groups in the sample, and hence, they will be used in the tailored weighting. As we set out the principles of the S-TWA in the introduction, ideally, variables that do not distinguish response from non-response in the selected sub-groups should be excluded from the creation of the tailored weights for these sub-groups regardless of whether or not they predict response in other sub-groups. It may very well be argued that such variables should be kept in the model if they predict response in other sub-groups even if they are not important predictors for the sub-groups under investigation. However, the resultant weights in this case may not be fully tailored to the sub-group in question, they are, to an extent, standard.

Model the response propensity separately for each sub-group: with this approach the response propensity can be modelled separately for each sub-group in the sample by restricting the modelling to the given sub-group (i.e. estimate separate weighting models for each sub-group). The subgroup-specific weighting models will only use variables that are associated with the response probability in the relevant sub-group. Accordingly, the set of weighting variables for a given sub-group may exclude variables from the SWA that do not predict response in the sub-group in question, and include the variables that are proposed for the S-TWA for this sub-group. The weights will then be calculated, separately for each sub-group, as the product of the inversed predicted probabilities from the sub-group estimated models. Thus, applying this approach will result in a subset of tailored weights for each sub-group. However, the resultant subsets of tailored weights can be combined to form an overall set of *TWs.*

It may be likely that the two approaches yield similar results. However, we tend to promote the second approach, especially if many categorical variables are suggested for the S-TWA since it will be more practical in this case, and also because it allows exclusion of the variables that are not significant at the sub-group level.

In this investigation we apply both methods of subgroup-tailored weighting as AWAs. This strategy should enable us to report on whether the two approaches can result in different outcomes. We refer to the tailored weights resulting from the *interaction-based approach* as TWs_1, whereas weights resulting from *modelling the response propensity separately for each sub-group* are denoted as TWs_2. The construction of the TWs_1 and TWs_2 is explained in the next section.

3.6 Construction of the tailored weights (TWs)

3.6.1 Interaction-based approach

To apply this approach we created two indicators. One of these is for retired sample members and the other is for those who were born in 1965 or after (1=retired, 0=non-retired; and 1=born in 1965 or after, 0=was not born in 1965 or after). We used the same weighting models of the SWA, and we added interactions of each indicator and its relevant proposed weighting variables introduced in section 3.4. The results of modelling the response propensity using this approach are presented in table (3.2). Note that we do not include 'age' and 'household with children' in these models as there are two variables used in the tailored weighting that can substitute for these ('born in 1965 or after and' 'has their own children' respectively).

We can already indicate that the results regarding the variables proposed for the S-TWA here are similar to those from modelling the response propensity separately for each sub-group which will be presented next. Thus, the effect of including these variables in the weighting process will be discussed in details in the next section. However, the major findings here will be highlighted.

First, with respect to the standard weighting variables (i.e. the variables used in the SWA), most variables have the same effect on the response propensity as in the SWA.

Second, none of the main effects of our new added variables appear to be significant in the models displayed in table 3.2 (with the exception of ‘has their own children’ as this substitutes for ‘household with children’). The significance of these variables is rather reflected in their interactions with the indicators of the two sub-groups in question. For example, the variables: religious and likes their current neighbourhood do not seem to be significant in predicting response for the sample as a whole. However, the interactions of these variables with retired sample members and those who were born in 1965 or after, respectively, appear to be significant suggesting that these variables are important in predicting response in the sub-groups under investigation.

This finding confirms our hypothesis that non-response process may be different in the selected sub-groups than in the sample in general. In addition, it shows that some of the factors responsible for non-response in these sub-groups are different than the factors responsible for non-response in the other subgroups in the sample. Furthermore, based on this finding, one can expect our proposed variables to be significant when modelling the response propensity separately for each sub-group as will be shown next.

Table 3.2 Response propensity models based on the AWA (interaction-based): modelling response in wave *t* conditional on responding in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Female	1.21*	1.28***	0.93	1.31**	1.12	1.27*	1.35**
White	1.46*	1.53***	1.30*	1.42*	0.88	1.51*	1.49**
Home owner	1.15*	1.19*	1.22*	1.31**	0.89	1.33**	1.13
Has GCE qualification or above	1.22*	0.92	0.87	1.08	1.26*	1.11	1.39**
Employed	0.85	1.09	1.06	1.23*	0.91	0.88	1.19*
Others present in interview	1.05	0.87	1.08*	1.04	1.39**	1.33*	1.07
Single-person household	0.93*	1.30	0.88**	1.10	0.96	0.76*	0.92
Living in a flat	0.96	1.14	0.89	0.83*	1.03	0.87*	0.92
Living in other type of house	0.88	0.81	0.69*	0.92	0.94	0.72*	1.05
1 or 2 persons in employment	1.05	0.93	0.92	0.89*	1.08	1.11	1.20
3 + persons in employment	0.91	0.68*	1.11	0.88	0.92	0.65*	0.93
South-East	0.94	0.93	1.36*	1.40*	0.77	1.07	0.85
South-West	1.29	0.95	1.19	1.17	1.20	0.88	1.05
East Anglia	0.91	0.93	1.66*	1.51*	0.89	1.03	1.33
The Midlands	0.97	0.89	1.22	1.26	0.86	1.11	0.91
The North	1.03	0.87*	1.34*	1.22	0.81	1.41*	1.19
Wales	0.93	0.74*	1.58*	1.02	0.67*	0.95	0.86
Scotland	0.89	0.81	1.29	1.44*	0.31**	1.10	0.83*
Retired	1.21*	1.26	1.18*	0.92	0.53*	0.71*	1.06
Religious	0.84	1.05	1.11	1.15	0.87	1.01	1.19
Retired X religious	1.22*	1.15*	1.49	1.27*	1.68*	1.11	0.87
Has more energy than average at their age	1.13	1.09	0.87	1.11	1.18	0.86	1.09
Has less energy than average at their age	0.92	1.06	1.12	1.11	0.94	0.90	0.89
Retired X has more energy than average at their age	1.11	0.97	1.59*	1.07	0.88	1.68*	1.39*

* The table is continued in the next page for the rest of the variables. The entries are odds ratios. In every wave response is modelled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. X indicates an interaction term.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.2 (continued)

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Retired X has less energy than average at their age	0.51***	0.39***	0.44**	0.38**	0.42**	0.78	0.91
Supports a political party	1.08	0.93	1.11	1.05	0.92	0.88	1.07
Retired X supports a political party	1.31*	1.30*	1.16	1.19	1.34*	1.33	1.08*
Financially okay	0.89	0.81	0.78	0.91	0.85	1.11	1.22
Financially struggling	0.88	0.93	0.82	0.87	0.77	1.09	0.90
Retired X financially okay	0.96	0.94	0.87	1.12	0.90	0.82*	1.16
Retired X financially struggling	1.10	0.85*	0.89*	1.24	0.84*	0.96	0.82*
Has a car	1.29	1.31	1.13	0.94	1.03	1.12	0.92
Retired X has a car	1.26*	1.19*	0.90	0.94	1.11*	0.71	1.09*
Was born in 1965 or after	0.81*	0.79*	1.03	1.16	1.36*	1.20	1.41*
Likes the current neighbourhood	1.22	1.28	1.10	1.31	1.22	1.30	1.02
Born in 1965 or after X likes the current neighbourhood	1.14*	0.73	1.10*	0.77	1.26	1.29*	0.89
Left school aged 14 or less	0.71	0.86	0.89	1.18	0.82	0.88	0.79
Left school aged 17 or over	1.27	1.22	1.30	1.05	0.89	1.18	0.77
Born in 1965 or after X left school aged 14 or less	1.15	0.81	0.73*	0.49*	0.91	0.78*	0.90
Born in 1965 or after X left school aged 17 or over	1.17	1.19	0.84*	1.18	0.88*	0.91*	1.22
Has their own children	0.92	1.25*	1.08	1.22*	0.97	1.19*	1.04
Born in 1965 or after X has their own children	1.10	1.23*	1.07	1.26*	1.06	1.28*	0.87
Born in 1965 or after X financially okay	0.92	0.89	1.02	1.12	1.11	1.18*	1.21*
Born in 1965 or after X financially struggling	1.15*	0.75	1.28*	0.93	1.34*	0.86	0.79
Born in 1965 or after X has a car	0.96	0.72*	0.83*	0.91	1.08	0.88*	1.21
N	9,593	8,699	8,218	7,863	7,496	7,152	6,878
Pseudo R^2	0.030	0.033	0.036	0.035	0.032	0.034	0.033

* The entries are odds ratios. In every wave response is modelled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. X indicates an interaction term. * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

For the set of responding sample members in the 8 waves, the TWs_1 at wave 8 were calculated as the product of the inversed predicted probabilities from the models in table (3.2), and wave 1 non-response/design weights as shown in equation (3.3).

$$TW_{1i} = D_i * \prod_{t=2}^{8} r_{ti}^{-1} \quad (3.3)$$

Where TW_{1i} is the *interaction-based* subgroup-tailored weight at wave 8 for respondent *i*; r_{ti} is the predicted probability for respondent *i* from wave *t* model (*t*= 2, 3,…, 8); *i*= 1,..., $n_{1,..,8}$ (where $n_{1,..,8}$ is the number of sample members who responded at every wave from 1 to 8); and D_i is wave 1 non-response/design weight for respondent *i*.

The distribution of the TWs_1 will be presented and discussed, together with the TWs_2 and the *SWs* in section 3.7.

3.6.2 Modelling the response propensity separately for each sub-group

To model the response propensity separately for each sub-group we estimated three different groups of weighting models. Recall that we have three sub-groups in the sample: retired sample members, sample members who were born in 1965 or after, and non-retired sample members who were born before 1965. However, the S-TWA focuses on the first two sub-groups. Thus, for each of these two sub-groups, the weighting models excluded some of the variables used in the SWA which are not important in the given sub-group in terms of predicting response, and included the relevant proposed weighting variables. This adjustment makes the sets of weighting variables used in the creation of the TWs_2 for each of the selected sub-groups different from each other and from the set of variables used in the SWA. As for the weighting models of the third sub-group (non-retired who were born before 1965), this used the same variables from the SWA.

The remainder of this section discusses, in detail, the results from modelling the response propensity separately for each sub-group. Our discussion here will be limited to the variables proposed for the SWA as the other variables result in similar results to the SWA.

Modelling response propensity for retired sample members

Some of the variables that were used in the SWA were dropped in this analysis, as described below. Also, new variables were added. The added variables are our proposed variables for the S-TWA for the retired sample members. Furthermore, the weighting models were estimated using only the set of retired sample members. The results of the weighting models of the retired sample members are presented in table 3.3.

Dropped variables

Employment status: employment status is an important factor that predicts response propensity in the analysis of non-response because it is a good predictor of the probability of contact. Normally, those who are in full-time employment are more difficult to contact since they are less likely to be at home (Groves and Couper, 1998). However, employment status was excluded from the set of weighting variables in this case as all of the sample members in this sub-group are retired.

Number in employment in household: in any survey that contacts sample members at their home, a successful contact attempt with any household depends on whether some (or at least one) of the household members are (is) actually at home to respond to the contact attempt. Thus, in this context, the number of household members in employment can be negatively associated with successful contact attempts. Consequently, households with more individuals in full-time employment are less likely to respond compared to households that have less number of individuals in full-time employment. This is confirmed by our results from the SWA in table 3.1. However, dealing with retired sample members guarantees that there is at least one household member who is not in a full-time employment and hence it is more likely to successfully establish contact in this case. Since this applies to all retirees, this variable was excluded from the choice of weighting variables for retired sample members.

Added variables (proposed for the tailored weighting of retired respondents)

Religion: religion was included in the model as a categorical variable with two categories: religious and non-religious (reference category). Most of the models in table 3.3 show that those who have a religion are more likely to respond than those who do not have a religion.

Respondent's energy compared to average at their age: this variable was included in the weighting models as a dummy variable with three categories: has the same energy as average at the same age (reference category), has more energy compared to the average at the same age and has less energy compared to the average at their age. As it can be seen from table 3.3, most of the models indicate that those who have more energy than average are more likely to respond than those who have the same energy as average. In contrast, sample members with less amount of energy compared to the average at their age are less likely to respond than those who have the same energy as average at their age. The explanation here is that 'energy' may be a strong indicator of the physical ability of a retired sample member to take part in the interview. Thus, retired individuals with more energy than average are likely to be in a good health condition, which may in turn increase the likelihood of successfully conducting the interview with older respondents. As for those who have less energy compared to people at their age, it is less likely that they will be cooperative compared to those with same energy as average.

Whether respondent supports a political party: this is a categorical variable with two categories: supports a political party and does not support a political party (reference category). As expected, our response propensity models in table 3.3 show that when this variable is significant, those who support a political party are more likely to respond than those who do not. This is in line with our hypothesis suggesting that retired sample members who have political views may feel more obligated to respond to the survey.

Subjective financial situation: the BHPS measures the subjective financial situation by asking respondents this question "how well would you say you yourself are managing financially these days?" In turn, respondents have to report their financial situation by selecting one of these options: living comfortably, doing alright, just about getting by, finding it quite difficult and finding it very difficult. Rearranging these options by combining the second option with the third, and the fourth option with the fifth, subjective financial situation was included in the models as a categorical variable with three categories: having a good financial situation (reference category), financially okay and financially struggling. The results suggest that, for retired respondents, those who are better off are more likely to respond than those who are less well off. The models indicate that both those who are financially okay and those who are financially struggling are less likely to respond than those with a good financial situation. These results are similar to the general findings of the effect of wealth on the response propensity.

However, recall that the evidence from the BHPS (for the whole sample) regarding financial situation is in contradiction with this finding (Uhrig, 2008). Thus, confirming our hypothesis, the results here indicate that the effect of financial situation on the response propensity is different for retired respondents than for the rest of the sample.

Having access to a car: this was included in the model as a categorical variable with two categories: has a car and has no car (reference category). Most of the models in table 3.3 show that retired respondents who have access to a car are more likely to maintain response than those who do not have access to a car. Our explanation for this is that, for retired sample members, having a car for personal use is indicative of a good physical health and relatively good financial situation.

Table 3.3 Response propensity models for retired respondents: modelling response in wave *t* conditional on responding in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Female	1.12**	1.19*	1.02	1.23*	1.40	1.38	1.19
White	1.31	1.68*	1.37*	1.13	1.26	1.64	1.59
Age	1.01*	0.99	1.25***	1.11***	1.18***	1.19***	1.25*
Age-squared	0.95*	0.98	0.99***	0.99**	0.99**	0.99***	0.99*
Home owner	1.30*	0.99	0.84	0.97	1.14*	1.47	0.84
Has GCE qualification or above	1.07	0.98	1.63*	1.20	1.41*	1.28	1.41
Others present in interview	0.88	1.18*	1.22*	1.14	1.41	1.13	1.78
Single-person household	0.99	0.97*	0.78**	1.17	0.64	0.73*	1.16
Household with children	1.32*	0.87	0.83	1.62*	0.73	0.79	1.11*
Living in a flat	0.60	0.73*	0.75	1.18	1.02	0.97	0.29
Living in other type of house	1.32	0.92*	0.86	0.70*	0.60	0.69*	0.66*
South-East	1.02	1.01	0.55	1.20	0.52	2.11	2.85
South-West	0.98	1.16*	1.12	1.39	0.63	1.27	1.93
East Anglia	1.41*	0.88	1.23*	1.35	0.67	1.51	1.20
The Midlands	1.32	0.73	1.01	1.43	0.46	1.77	1.82
The North	1.18	1.08	1.05	1.12	0.65	1.21	1.40
Wales	0.89	1.17	0.86	1.14	0.40*	1.42	1.24
Scotland	1.39	2.17	1.05	3.74*	0.50	0.96	0.87
Religious	1.03*	1.56*	1.34*	1.39*	1.16*	1.84	0.82
Has more energy compared to average at their age	0.92	1.06	1.24*	1.37	1.60	1.33*	1.46*
Has less energy compared to average at their age	0.46***	0.48***	0.49**	0.55**	0.53*	0.61	0.88
Supports a political party	1.08*	1.12*	0.86	0.94	1.10*	0.97	1.09*
Financially okay	0.90	1.10	1.13	1.38	1.39	0.89*	1.29
Financially struggling	1.04	0.87*	0.85*	0.93	0.88*	1.06	0.79*
Has a car	1.08*	1.16*	0.52	1.01	1.21*	0.65	1.20*
N	1,712	1,647	1,594	1,550	1,496	1,457	1,418
Pseudo R^2	0.037	0.038	0.039	0.035	0.038	0.036	0.036

* The entries are odds ratios. In every wave response is modelled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. The reference categories of the categorical independent variables in the table are male, non-white, not a home owner, does not have a GCE or above degree, others not present when interviewed, multi-person HH, household with no children, living in a house, London, non-religious, has the same energy as average as their age, does not support a political party, having good financial situation and does not have a car * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

For the set of responding retired sample members in the 8 waves, the tailored weights at wave 8 were calculated as the product of the inversed predicted probabilities from the models in table (3.3), and wave 1 non-response/design weights as shown in equation (3.4).

$$TW_{RRi} = D_i * \prod_{t=2}^{8} r_{ti}^{-1} \quad (3.4)$$

Where TW_{RRi} is the tailored weight at wave 8 for retired respondent i, based on modelling the response propensity separately for retired sample members; r_{ti} is the predicted probability for retired respondent i from wave t model (t= 2, 3,…, 8); i= 1,..., $n_{1,..,8}$ (where $n_{1,..,8}$ is the number of retired sample members who responded at every wave from 1 to 8); and D_i is wave 1 non-response/design weight for retired respondent i.

Modelling response propensity for those who were born in 1965 or after

Similar to modelling the response propensity for retired respondents, the weighting models for those who were born in 1965 or after were estimated by changing some of the weighting variables used in the SWA and by using the set of sample members who were born in 1965 or after. The results of the weighting models of those who were born in 1965 or after are displayed in table 3.4.

Dropped variables

Age: age is an important factor in predicting non-response. The literature indicates that, in general, elderly people are more likely to refuse to participate in the survey than younger respondents (Groves and Couper, 1998; Lepkowski and Couper, 2002). However, other research suggests that the youngest respondents in the sample are more difficult to locate as they have a higher tendency to move house, and even if they are located, they are still difficult to contact because they are less likely to be at home (Stoop, 2005). This pattern is very common among the vast majority of younger sample members. In this research, respondents who were born in 1965 or after fell into the age group 16-26 by the time the first wave of BHPS was conducted. This age group forms the youngest age group in the sample. However, preliminary analysis for this age group showed that age is not an important factor to predicting non-response within this age group. Thus, the weighting models for those who were born in 1965 or after were estimated without including the variable age.

Whether children in household: this variable was used to estimate the weighting models in the SWA. It indicates if there are children within the household. This is regardless of whether these children are the respondent's own children (i.e. could be nephews, nieces, etc…). Non-response theory suggests that the presence of children in the household is associated with high levels of response. This is because the presence of children in the household indicates more social integration (e.g. taking the kids to school or nursery) and hence it is easier to locate and contact households with children than single-person households or households with no children (Groves and Couper, 1998; Uhrig, 2008). However, one of the proposed weighting variables for those who were born in 1965 or measures the respondent's own children in the household. This variable somewhat substitutes for the presence of children in the household and therefore the latter was excluded from the weighting model of those who were born in 1965 or after.

Added variables (proposed for the tailored weighting of those born in 1965 or after)

Liking the neighbourhood: Liking the neighbourhood was included in the models as a categorical variable with two categories: likes their current neighbourhood and does not like their current neighbourhood (reference category). As shown in table 3.4, when this variable is significant, it indicates that those who like living in their neighbourhood are more likely to respond than those who do not like living in their neighbourhood. This result indicates that one's attachment to the neighbourhood where they reside may be particularly important in predicting response for those who were born in 1965 or after. In general, individuals who are not attached to their residence neighbourhood are likely to move house and hence may be difficult to track and re-establish contact with. However, this is especially more likely for younger sample members (those who were born in 1965 or after in our case) who are usually more mobile compared to their older counterparts.

School leaving age: To measure this variable, BHPS sample members were asked the following question: "how old were you when you left school". In return, if not still at school, respondents reported the age at which they left school. The reported ages range between 9 and 22. These answers were categorised into three categories: left school aged 14 or below, left school aged 15 or 16 (reference category) and left school aged 17 or above. At the time of wave 1, there was a small number of respondents who were still in school. This group of sample members does not allow valid estimation of the weighting models if they are treated as a separate category. This is especially the case in the weighting models after wave 2 as more cases from this category leave school as time goes on. Thus, these cases were classified with the category 'left school aged 17 or above' (since everyone in our sample aged 16+ at wave 1, eventually those who were still in school at the time of wave 1 will have left school aged 17+). Most of our models here suggest that both those who left school aged 14 or below and those who left school aged 17 or above are less likely to respond than those who left school aged 15 or 16.

Having children: In the BHPS data set there is a variable that refers to the number of the respondent's own children in the household. The value of this variable ranges from between 0 and 9. This variable was used to indicate whether the respondent has children or not. It was categorised into two categories: has their own children in household (by combining the numbers from 1 to 9 in one category) and does not have their own children in household (reference category). As expected, the results suggest that those who have their own children within the household are more likely to respond than those who do not have children in the household.

Subjective financial situation: similar to modelling the response propensity for retired sample members, financial situation here was included in the models as a categorical variable with three categories: having a good financial situation (reference category), financially okay and financially struggling. Unlike the findings for retired sample members, the evidence here suggests that those who are less well off are more likely to respond than those who are better-off. This result confirms that financial situation is indeed an important factor for predicting response for both retired sample members and those who were born in 1965 or after. However, and more importantly, it shows that the effect of this variable is different for the two sub-groups. Thus, a weighting strategy like the SWA which might not recognise this as it estimates its weighting models by assuming that the effect of such variable is similar for all sub-groups may result in a set of weights that does not properly adjust for non-response in estimates of financial phenomena which are related to the sub-groups in question.

Having access to car: Having access to a car was included in the model as a categorical variable with two categories: has a car and has no car (reference category). The results for this variable indicate that those who have a car for personal transport are less likely to respond that those who do not have a car. One possible explanation for this is that, for younger sample members (those who were born in 1965 or after), having a car may be a factor that stimulates the 'not at home pattern'. Thus, young sample members who have a car may be less likely to be contacted successfully than those who do not have a car.

Table 3.4 Response propensity models for those who were born 1965 or after: modelling response in wave *t* conditional on responding in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Female	1.09	1.36*	1.66	1.33	1.20*	1.95	1.25*
White	1.27**	1.18*	1.22*	1.07	0.58	1.24	1.17*
Home owner	1.12*	0.95	1.21	1.17	1.71	1.39**	1.57
Has GCE degree or more	1.28	1.08*	1.49	1.19*	1.29	1.63	1.46
Employed	0.71	0.94	1.07*	1.15*	1.37	1.13*	1.17
Others present in interview	1.28	1.45	1.47	1.54*	0.83	1.14*	0.99
Single-person household	0.80*	1.07	1.21	0.98	0.63*	0.87	0.69*
Living in a flat	0.79	0.65	0.65*	0.59	1.32	0.72*	1.83
Living in other type of house	0.84	0.66	0.55*	1.53	1.39	0.90	1.29
1 or 2 persons in employment	1.20	1.01	1.11	1.10	0.64*	0.69	0.77*
3 + persons in employment	0.90	0.53*	1.27	0.99	0.43*	0.78	1.93
South-East	0.69	1.15	2.31	0.89	1.53	0.61	0.37
South-West	1.13*	1.45	2.00	1.08	3.23	0.42	0.81
East Anglia	1.74*	1.21	1.19	1.25	1.07	2.37*	1.46*
The Midlands	1.05	1.21	1.16	0.69	1.67	0.70	0.64
The North	1.38	1.02	2.53	0.91	2.37	0.80	0.69
Wales	0.70	0.83	2.22	0.48*	1.30	0.53	0.39
Scotland	1.20	1.12	1.55	0.62	1.06	0.54	0.32
Likes their current neighbourhood	1.20*	0.98	1.42*	0.69	0.89	1.56*	0.97
Left school aged 14 or less	0.91	0.56	0.64*	0.40*	0.51	0.53*	0.49
Left school aged 17 or above	0.89	0.84	0.62*	1.04	0.77*	0.76*	1.06
Has their own children	0.86	1.18*	1.52	1.55*	0.97	1.63*	1.50
Financially okay	0.89	0.97	0.93	1.31	1.48	1.22*	1.11*
Financially struggling	0.84	0.83	1.14*	0.88	1.21*	1.25	0.90
Has a car	1.15	0.69*	0.86*	1.04	1.05	0.79*	1.18
N	1,933	1,862	1,798	1,757	1,695	1,651	1,576
Pseudo R^2	0.030	0.034	0.035	0.038	0.036	0.039	0.035

*The entries are odds ratios. In every wave response is modelled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. The reference categories of the categorical independent variables in the table are male, non-white, not a home owner, does not have a GCE or above degree, unemployed, others not present when interviewed, multi-person HH, living in a house, no one in employment in HH, London, does not like their current neighbourhood, left school aged 15 or 16, does not have their own children, having good financial situation and does not have a car * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

For sample members who were born in 1965 or after and who responded in the 8 waves, the tailored weights at wave 8 were calculated as the product of the inversed predicted probabilities from the models in table (3.4), and wave 1 non-response/design weights as shown in equation (3.5).

$$TW_{1965i} = D_i * \prod_{t=2}^{8} r_{ti}^{-1} \tag{3.5}$$

Where TW_{1965i} is the tailored weight at wave 8 for respondent *i* (who was born in 1965 or after) based on modelling the response propensity separately for sample members who were born in 1965 or after; r_{ti} is the predicted probability for respondent *i* from wave *t* model (*t*= 2, 3,…, 8); *i*= 1,..., $n_{1,..,8}$ (where $n_{1,..,8}$ is the number of those were born in 1965 or after who responded at every wave from 1 to 8); and D_i is wave 1 non-response/design weight for respondent *i*.

Modelling response propensity for those who are non-retired and born before 1965

The set of weighting variables used to estimate the weighting models for those who are non-retired and born before 1965 is the same as the set of weighting variables used in the SWA. However the models were only restricted to those who are non-retired and were born before 1965. Table 3.5 shows the results of modelling the response propensity in the 8 waves for this part of the sample. As expected, the results here are similar to the ones from the SWA. Overall, the results indicate that response is higher amongst females, white sample members, those with more education, employed individuals and members of multi-person households or households with children.

Table 3.5 Response propensity models for non-retired respondents who were born before 1965: modelling response in wave *t* conditional on responding in all of the previous waves.

	Wave2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Female	1.14	1.38**	1.14	1.19*	1.13	0.98	1.53*
White	1.29**	1.87**	1.44*	1.67*	0.70	1.83	1.33**
Age	1.03*	1.08***	1.07**	1.14***	1.10***	1.17***	1.24***
Age-squared	0.99*	0.99***	0.99***	0.99***	0.99***	0.99***	0.99***
Home owner	1.04	1.50**	1.38*	1.57**	1.13*	1.63*	1.21
Has GCE degree or more	1.25**	1.11	0.96	0.98	0.91	1.01	1.22*
Employed	1.07	1.02	1.25*	1.14	1.17*	1.08	1.20*
Others present in interview	1.08*	1.12	0.86	1.05	1.07	1.51*	1.49
Single-person household	1.29	0.91*	0.90	0.92	1.01	0.72*	0.83*
Household with children	1.25**	1.06	1.30	1.32	1.02	1.05	1.26*
Living in a flat	0.72	0.79*	1.09	0.87	0.77*	0.83	0.79
Living in other type of house	1.05	1.28	0.46*	0.89	1.49	0.65*	1.64
1 or 2 persons in employment	1.01	0.65*	0.93	0.96	1.25	1.08	0.71
3 + persons in employment	0.90	0.56*	0.81	0.61	0.58	0.34*	0.42*
South-East	1.09	0.80	1.37	1.75	1.41	1.19	0.80
South-West	1.18	0.78	1.12	1.32	1.58	0.91	0.88
East Anglia	0.92	0.84	2.05*	2.71*	0.89	1.34	1.17
The Midlands	0.92	0.72	1.25	1.42*	1.22	1.15	0.82
The North	1.31*	0.60	1.37	1.26	0.86	2.08**	0.89
Wales	0.90	0.56	1.36	1.34	1.70*	1.59	1.20
Scotland	0.84	0.45*	1.78	1.74	0.50	1.25	0.44*
N	5,948	5,190	4,826	4,556	4,305	4,044	3,884
Pseudo R^2	0.029	0.033	0.034	0.033	0.032	0.034	0.034

*The entries are odds ratios. In every wave response is modelled conditional on responding in all of the previous waves. The model in a given wave used variables from the previous wave. The reference categories of the categorical independent variables in the table are male, non-white, not a home owner, does not have a GCE or above degree, unemployed, others not present when interviewed, multi-person HH, household with no children, living in a house, no one is in employment in HH and London * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

For sample members who are non-retired and were born before 1965 (remaining sample), and who responded in the 8 waves, the tailored weights at wave 8 were calculated as the product of the inversed predicted probabilities from the models in table (3.5), and wave 1 non-response/design weights as shown in equation (3.6).

$$TW_{RSi} = D_i * \prod_{t=2}^{8} r_{ti}^{-1} \tag{3.6}$$

Where TW_{2i} is the tailored weight at wave 8 for respondent *i* (who is non-retired and was born before 1965) based on modelling the response propensity separately for sample members who are non-retired and were born before 1965; r_{ti} is the predicted probability for respondent *i* from wave *t* model (*t*= 2, 3,…, 8); *i*= 1,..., $n_{1,..,8}$ (where $n_{1,..,8}$ is the number of those who are non-retired and born before 1965 who responded at every wave from 1 to 8); and D_i is wave 1 non-response/design weight for respondent *i*.

Since the three sub-groups in the analysis are non-overlapping, the sub-sets of tailored weights resulting from modelling the response propensity for each sub-group were then put together to form our second set of tailored weights (TWs_2) as follows:

$$TWs_2 = TW_{RR} \cup TW_{1965} \cup TW_{RS} \tag{3.7}$$

The distribution of the SWs, TWs_1 and TWs_2

In this section we discuss and present the distribution of the weights resulting from the SWA (*SWs*) and the two methods of the S-TWA (TWs_1 and TWs_2). Table (3.6) presents the measures of central tendency and dispersion for the three sets of weights. For each set of weights, these statistics are presented separately for retired respondents, those who were born in 1965 or after and for all respondents.

By looking at the standard errors of the three sets of weights under investigation, we can notice that these weights have very similar dispersion within all sets of respondents. This is confirmed with the coefficients of variations (CV) which are almost identical for the three sets of weights across the three groups of respondents. Thus, with the same amount of variation in all sets of weights, it seems reasonable to expect rather similar results in terms of precision for equivalent estimates constructed with *SWs*, TWs_1 or TWs_2.

As for the average weight value, this appears to show different results. While TWs_1 and TWs_2 have the same average weight values across the three sets of respondents, *SWs* seem to have smaller weights sizes on average compared to the tailored sets of weights. This is also the case with the medians, and the first and third quintiles values indicating that, for most cases in the sample, TWs_1 and TWs_2 contains fairly larger weights compared to the *SWs*. These results suggest that the S-TWA resulted in somewhat different weights than the SWA in terms of the average weights value. Accordingly, we expect this to affect the magnitude of some of the estimates resulting from the S-TWA, possibly to an extent that makes them significantly different than their equivalent estimates resulting from the SWA.

Additionally, we do not expect to find considerable differences between estimates resulting from TWs_1 and TWs_2 as the distributions of these two sets are very similar both in terms of dispersion and average weights value. This, in turn, suggests that our two approaches of sub-group tailored weighting may have similar effects on the resultant weights.

Table 3.6 The distribution of SWs, TWs_1 and TWs_2.

	Retired respondents			Born 1965 or after			The whole sample		
	SWs	TWs_1	TWs_2	SWs	TWs_1	TWs_2	SWs	TWs_1	TWs_2
Std.dev	0.66	0.68	0.69	0.67	0.68	0.69	0.63	0.65	0.66
Mean	1.98	2.12	2.10	1.79	1.86	1.87	1.58	1.65	1.64
CV	0.33	0.32	0.33	0.37	0.37	0.37	0.40	0.39	0.40
Min	0.49	0.54	0.59	0.40	0.46	0.44	0.32	0.35	0.33
Q1	1.61	1.71	1.73	1.42	1.58	1.55	1.26	1.31	1.36
Median	1.80	1.89	1.87	1.69	1.76	1.77	1.49	1.54	1.55
Q3	2.11	2.35	2.30	2.06	2.12	2.15	1.86	1.91	1.93
Max	6.37	10.86	10.05	6.06	7.24	7.75	6.88	11.16	11.89

*CV is the coefficient of variation (CV=Std.dev/Mean).

3.7 Analysis and results

At this stage, there are three different sets of weights in our analysis (SWs, TWs_1 and TWs_2). Each set of weights is designed to adjust for cumulative non-response between waves 1 and 8. This means that the weights from each set are available for sample members who responded in all of the first 8 waves of the survey. Thus, for our substantive analysis we used a balance panel of those who responded in all waves from wave 1 up to and including wave 8 (6,753 respondents). In this sample, there are 1,402 retired respondents and 1,525 sample members who were born in 1965 or after.

Our investigation focuses on whether TWs_1 and TWs_2 affect estimates produced from the sub-groups under investigation (retired respondents and those who were born in 1965 or after) and estimates based on the whole sample differently compared to the *SWs.* In other words, we investigate whether our proposed S-TWA adjust for non-response differently as opposed to the SWA.

To examine this, we carried out two sets of analyses. The first is concerned with retired respondents. In this analysis we estimate a model to investigate the determinants of psychological well-being for retired respondents. Psychological health is an important social aspect which is known to be affected, either positively or negatively, by later life transitions such as retirement (Kim and Moen, 2002). Thus, it might be appropriate to evaluate our weights by conducting this analysis on retired respondents. Since the set of weights resulting from the S-TWA contains weights for respondents from two major subgroups in the sample, we also test the effect of the S-TWA on full sample estimates. Therefore, we fit another model for the determinants of psychological well-being using the full sample. In both analyses (retired respondents and the full sample) we estimate the same model by using the *SWs*, TWs_1 and TWs_2 separately.

The second set of analyses is concerned with those who were born in 1965 or after. In this part of our analysis we estimate a model to investigate the determinants of the desire for residential mobility (DRM). DRM is a social phenomenon that expresses individuals wish to change their address (Sadig and Banany, 2015). Since residential mobility is more common amongst younger individuals, it may be of interest to assess the S-TWA and the SWA by investigating DRM for those who were born in 1965 or after (the younger respondents in our sample). Also, and similar to the analysis of retired respondents, we investigate the determinants of DRM for those who were born in 1965 or after and for the whole sample separately. Yet again, in each analysis, we estimate the same model by using the *SWs*, TWs_1 and TWs_1 separately.

Accordingly, there will be 12 substantive models in this investigation. Six of these are for the analysis of psychological well-being and the other six are for the analysis of DRM. For simplicity, we can identify these models as follows:

Psychological well-being

Model 1 is estimated using retired respondents and the *SWs.*

Model 2 is estimated using retired respondents and the TWs_1.

Model 3 is estimated using retired respondents and the TWs_2.

Model 4 is estimated using the full sample and the *SWs.*

Model 5 is estimated using the full sample and the TWs_1.

Model 6 is estimated using the full sample and the TWs_2.

DRM

Model 7 is estimated using those who were born in 1965 or after and the *SWs.*

Model 8 is estimated using those who were born in 1965 or after and the TWs_1.

Model 9 is estimated using those who were born in 1965 or after and the TWs_2.

Model 10 is estimated using the full sample and the *SWs.*

Model 11 is estimated using the full sample and the TWs_1.

Model 12 is estimated using the full sample and the TWs_2.

This strategy allows a fair comparison across the three sets of weights under investigation on different sets of respondents. This is because the same model will be estimated with each set of weights separately. Thus, differences between the equivalent estimates resulting from the application of different sets of weights can be said to be due differences between the weights as the estimation method is held constant. Consequently, this enables one to report on differences between the SWA and the S-TWA, but it also permits comparisons between the two different approaches that we used to create the tailored weights (interaction-based approach and modelling response propensity separately for each sub-group).

3.7.1 Psychological well-being

Measure of psychological well-being

There is a range of variables that measures psychological well-being in the BHPS. But, the most appropriate variables are probably the ones that are available within the General Health Questionnaire (GHQ). This is because the GHQ variables are reliable measures of psychological well-being (Taylor, Jenkins and Sacker, 2011). These are 12 items and they are obtained by asking the following questions:

- Have you recently been able to concentrate on whatever you're doing?

- Have you recently lost much sleep over worry? *

- Have you recently felt that you were playing a useful part in things?

- Have you recently felt capable of making decisions about things?

- Have you recently felt constantly under strain? *

- Have you recently felt you couldn't overcome your difficulties? *

- Have you recently been able to enjoy your normal day-to-day activities?

- Have you recently been able to face up to problems?

- Have you recently been feeling unhappy or depressed? *

- Have you recently been losing confidence in yourself? *

- Have you recently been thinking of yourself as a worthless person? *

- Have you recently been feeling reasonably happy, all things considered?

Respondents are asked to rate each item on a four-point scale: better than usual, same as usual, less than usual and much less than usual. The codes assigned to each answer are 0, 1, 2 and 3 respectively. Questions marked as * are coded in reverse. The GHQ items are added together to construct a general score which measures the mental distress of the cases in the sample. This score is known as the likert score (or likert scale). The likert score ranges from 0 to 36. Low scores indicate high feelings of well-being; meanwhile, high scores indicate high stress. The likert score was used in this analysis as the measure of psychological well-being (dependent variable).

It should be noted that the GHQ in the BHPS is a self-completion questionnaire. Therefore, it is likely that those who complete such questionnaires are in relatively good health status. This may not necessarily be the case for all retired respondents (the older respondents in the analysis sample) which may affect the estimates resulting from this analysis. However, this issue applies across the three sets of weights that are being evaluated. Thus, by holding the modelling approach constant and varying the weights, differences between the resultant estimates will be solely due to differences across the weighting schemes. Therefore, this analysis helps achieving the objective of this investigation.

To test the effect of the S-TWA on descriptive statistics, we categorised the likert score into two categories indicating good psychological health and bad psychological health[20]. We calculated the proportion of retired sample members in each category using SWs, TWs_1 and TWs_2 separately for a simple comparison. The resultant proportions and the associated standard errors are presented in table 3.7. As can be seen from the table, the standard errors of all proportions are almost identical indicating no difference in terms of impact on estimates precision levels across the three sets of weights. However, while TWs_1 and TWs_2 resulted in similar proportions, SWs resulted in a slightly different proportion (with a difference of 1%). This result indicates that the S-TWA may have a different impact on the magnitude of the estimate as compared to the SWA.

Table 3.7 Proportions of retired sample members with good and bad psychological health.

	Using SWs		Using TWs_1		Using TWs_2	
	%	SE	%	SE	%	SE
Good Ps.health	86	.0029	87	.0028	87	.0029
Bad Ps.health	14	.0029	13	.0028	13	.0029

* Ps.health is psychological health. SE is the standard error.

[20] Values from 0 to 18 represent good psychological health while values from 19 to 36 indicate bad psychological health.

Modelling psychological well-being

The outcome variable in this analysis (likert score) is a continuous variable. Similar to chapter 1 and 2, the structure of the data (multiple observations per person) allows the application of panel data models. Thus, we estimated a random effects OLS regression model to investigate the determinants of psychological well-being for retired respondents and the whole sample separately. As explained earlier, each model was estimated three times by varying the weights between SWs, TWs_1 and TWs_2. Psychological well-being is known to be associated with measures of ethnicity, age, cohabitation, wealth and health, (Taylor, Jenkins and Sacker, 2011; Kohler, Behrman and Skytthe, 2005; Ryan and Frederic, 2006). Thus, the variables used to model psychological well-being were selected to correspond to these measures. These variables are: Race, age, whether respondent lives with a partner, savings, health status and income. Additionally, other variables such as time and gender were also included in the model for control.

Before discussing the results from modelling psychological well-being, we first set out our criterion for identifying differences between estimates in both of our substantive analyses (psychological well-being and desire for residential mobility). We used the same hypotheses testing methods (using confidence intervals) that we used in the substantive analysis in chapter two to identify significant differences between equivalent estimates adjusted with different weights. For convenience, we re-explain this approach below.

To identify significant differences between equivalent coefficients estimated with the different weights, we conduct hypotheses testing on the differences between equivalent estimates adjusted with the SWA and S-TWA using 95% Confidence Intervals (CI). Our test involves two essential steps. The first step is to construct 95% CIs of the difference between each two equivalent coefficients that are adjusted with the SWA and S-TWA. Such CIs specify the range of values within which the difference between each two equivalent coefficients may lie. For example, if β_{SW}, β_{TW1} and β_{TW2} denote a given set of equivalent population parameters estimated by the equivalent set of coefficients b_{SW}, b_{TW1} and b_{TW2} which are adjusted with *SWs*, TWs_1 and TWs_2 respectively, we construct two CIs to test whether b_{SW} is different than b_{TW1} and b_{TW2}. These are CIs for (b_{SW} - b_{TW1}) and (b_{SW} - b_{TW2}). All CIs are 95% CIs, and are given by 3.8 below:

$$(b_{SW} - b_{TWi}) \pm 1.96 * S_{b_{SW}-b_{TW}} \tag{3.8}$$

Where $S_{b_{SW}-b_{TWi}}$ is the standard error of (b_{SW} - b_{TWi}) and is given by 3.9 below; and *i*=1,2.

$$S_{b_{SW}-b_{TWi}} = \sqrt{S^2(b_{SW}) + S^2(b_{TWi}) - 2 * Cov(b_{SW}, b_{TWi})} \tag{3.9}$$

Where $S^2(b_{SW})$ and $S^2(b_{TWi})$ are the variances of $b_{SW}\ and\ b_{TWi}$ respectively; $Cov(b_{SW}, b_{TWi})$ is the covariance of $b_{SW}\ and\ b_{TWi}$; and *i*=1,2.

The second step is to use the constructed CIs to test whether there is a significant difference between each two equivalent coefficients adjusted with the SWA and S-TWA (i.e. is there a significant difference between $b_{SW}\ and\ b_{TWi}$?). That is to test the following hypothesis:

H_0: β_{SW} - β_{TWi} =0 against H_a: β_{SW} - β_{TWi} ≠0; *i=1,2.*

Note that H_0 is rejected (i.e. there is a significant difference between b_{SW} and b_{TWi}, i=1,2) if the relevant CI does not include 0.

We apply the same test in both of our substantive analysis (psychological well-being and desire for residential mobility), and we present all 95% CIs for the difference between each two equivalent estimates (CI of [b_{SW} - b_{TW1}] and CI of [b_{SW} - b_{TW2}]) in the relevant results table.

Table 3.8 and table 3.9 present the results from modelling psychological well-being for retired respondents and the full sample respectively. Starting with the results from the retired respondents models (in tables 3.8), it can be seen that all equivalent estimates across the three models have the same significance level. This result is consistent with the earlier observation that the weights have similar distributions in terms of dispersion.

Additionally, we can immediately notice, as anticipated, that TWs_1 and TWs_2 result in very similar estimates. Most of the coefficients resulting from these two sets of weights are approximately equal. As for the coefficients resulting from the *SWs,* overall, these estimates are also similar to the estimates resulting from TWs_1 and TWs_2. However, the coefficient of 'has a good health condition' which is produced through the *SWs* appears to be significantly different from its equivalent coefficients estimated with TWs_1 and TWs_2. The difference is indicated by the two CIs of the difference between (b_{SW} - b_{TW1}) and (b_{SW} - b_{TW2}) respectively, where b here is the coefficient of 'has a good health condition'. Both CIs do not include 0 indicating that the estimate in question (adjusted with *SWs*) is significantly different than its equivalent estimates adjusted with TWs_1 and TWs_2. Thus, as suggested by this result, the S-TWA and the SWA may result in significantly different results with respect to estimates constructed from the sub-group selected for the tailored weighting.

Focussing on the models for the whole sample (in table 3.9), the results here do not show evidence of significant differences between the estimates resulting from the SWs and their equivalent estimates constructed with TWs_1 and TWs_2. All CIs of the difference between (b_{SW} - b_{TW1}) and (b_{SW} - b_{TW2}) include a zero. However, we can still notice that the coefficient of 'has a good health condition' in the model estimated with SWs is rather different compared to its equivalent coefficients in the models estimated with TWs_1 and TWs_2. Although our CI test here does not suggest that this difference is significant, such differences may matter in the interpretation of the results in some substantial analyses, and may play important role in modifying our understanding of some social phenomena. Thus, these differences are not trivial.

Table 3.8 Random effects OLS regression models of the determinants of psychological well-being for retired respondents.

	Using SWs	**Using TWs_1**	**95% CI of (b_{SW} - b_{TW1})**		**UsingTWs_2**	**95% CI of (b_{SW} - b_{TW2})**	
Years 1995 to 1998	0.131	0.030	-0.147	0.349	0.084	-0.202	0.296
Female	0.841**	0.884**	-0.351	0.256	0.861**	-0.626	0.586
White	-0.979	-1.268	-2.245	2.832	-1.766	-2.057	3.631
Age	0.020	0.022	-0.020	0.016	0.022	-0.038	0.034
Living with a partner	-1.027***	-1.042***	-0.452	0.482	-1.103***	-0.392	0.544
Has savings	-0.135**	-0.335**	-0.073	0.473	-0.322**	-0.086	0.460
Has a good health condition	-0.726*** [a]	-1.607*** [a]	0.321	1.441	-1.605*** [a]	0.313	1.445
Income/1000	-0.003	-0.004	-0.042	0.026	-0.004	-0.024	0.026
N	1,402	1,402			1,402		
σ	2.79	2.89			2.89		
ρ	0.46	0.47			0.47		

*All models are estimated by using a balanced panel of retired sample members who responded in the first 8 waves. The reference categories of the independent variables are: years 1991 to 1994, male, non-white, does not live with a partner, has no savings and has a bad health condition. [a] indicates a significant difference between the equivalent estimates adjusted with the *SWs* and both sets of *TWs*. * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

Table 3.9 Random effects OLS regression models of the determinants of psychological well-being for the whole sample.

	Using SWs	Using TWs_1	95% CI of (b_{SW} - b_{TW1})		Using TWs_2	95% CI of (b_{SW} - b_{TW2})	
Years 1995 to 1998	0.262	0.213	-0.067	0.165	0.248	- 0.089	0.117
Female	1.217***	1.233***	-0.252	0.220	1.226***	-0.245	0.227
White	-0.580**	-0.575**	-0.519	0.509	-0.631**	-0.467	0.569
Age	0.003*	0.006*	-0.009	0.003	0.005*	-0.008	0.004
Living with a partner	-0.482***	-0.485***	-0.180	0.186	-0.504***	-0.161	0.205
Has savings	-0.225***	-0.223***	-0.127	0.123	-0.228***	-0.123	0.129
Has a good health condition	-0.955***	-1.141***	-0.453	0.825	-1.162***	-0.394	0.808
Income/1000	-0.001	-0.002	-0.006	0.008	-0.002	-0.005	0.007
N	6,753	6,753			6,753		
σ	2.92	2.93			2.93		
ρ	0.36	0.37			0.37		

*All models are estimated by using a balanced panel of those who responded in the first 8 waves. The reference categories of the independent variables are: years 1991 to 1994, male, non-white, does not live with a partner, has no savings and has a bad health condition. * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

3.7.2 Desire for residential mobility (DRM)

Measure of DRM

It might not be common for social surveys to include a direct question about whether respondents have a desire to change their address. However, in the BHPS, respondents are asked every year if they would prefer to move house. This item was used here as our outcome variable. Accordingly, if respondents report a preference for moving house, this was taken as an indication of a desire for residential mobility. Thus, the dependent variable in this part of the analysis was a binary variable, indicating whether respondents have DRM or not. This variable is identified by equation 3.8.

$$DRM_i = \begin{cases} 1, & if\ respondent\ i\ has\ intention\ to\ move\ house \\ 0, & otherwise \end{cases} \tag{3.8}$$

Modelling DRM

To model DRM we used a random effects logistic regression. This was done for those who were born in 1965 or after and for the whole sample separately. For each set of respondents, the model was estimated, with the three sets of weights under investigation. The independent variables used in this analysis are gender, race, age, household size, number of rooms in the accommodation, possession of savings and housing tenure. Our choice for these variables was inspired by the literature of residential mobility (e.g. Sanbonmatsu *et al*, 2011) and the availability of these variables across the 8 waves used in this investigation.

The results from modelling DRM are displayed in table 3.10 for those who were born in 1965 or after, and in table 3.11 for the full sample. Both tables present odds ratios. As was done in the analysis of psychological well-being, we will use 95% CIs of the differences between estimates constructed with *SWs* (b_{SW}) and their equivalent estimates constructed with TWs_1 (b_{TW1}) and TWs_2 (b_{TW2}) in turn to test if the SWA and the two S-TWA result in different estimates

Focussing on the models for those who were born in 1965 or after (in table 3.10) first, we can see that, in general, the resultant estimates are similar across the three models. Aside from having the same significance levels, most coefficients are similar in terms of magnitude. This is more so between estimates resulting from TWs_1 and TWs_2 than between estimates resulting from *SWs* and either of the two tailored weights.

However, *SWs* produced one major difference compared to TWs_1 and TWs_2. Namely, the coefficient of 'member of a large household' appeared to be significantly different than its equivalent coefficients estimated with TWs_1 and TWs_2. This is confirmed by the by the two CIs of the difference between this estimate and its two equivalent estimates which are constructed with TWs_1 and TWs_2. Both CIs do not contain 0. This result is in line with the results from the retired respondents' models suggesting that the S-TWA may indeed result in some differences in comparison with the SWA, especially when the analysis is restricted to the sub-groups used to create the tailored weights.

Turning to models concerning the whole sample (in table 3.11), the results here do not show significant differences across estimates. By reference to our 95% CIs test, it can be seen that all the CIs of the differences between estimates adjusted SWs and their equivalent estimates adjusted with TWs_1 and TWs_2 respectively include a zero value of the difference indicating that each two equivalent estimates are not significantly different. However, and similar to the analysis of psychological well-being, we can still see a difference between the coefficient of ‘member of a large household’ in the model estimated with *SWs* and its equivalent coefficients in the models estimated with TWs_1 and TWs_2. As discussed before, such differences might also be important, and they indicate that the S-TWA may also affect total sample estimates. Based on this result, it may be reasonable to expect significant differences between total sample estimates resulting from the S-TWA and the SWA if other sub-groups are considered for the sub-group tailored weighting.

To sum up, based on this investigation, the results suggest that the two approaches of weighting (SWA and S-TWA) are similar in their overall effect on estimates. However, the S-TWA may have a different impact on some estimates. Based on our CIs tests, these differences were proved to be significant. Assuming that the weighting models estimated in the S-TWA express the response process in the sub-groups under investigation better than the weighting models of the SWA, estimates that turned out to be different with the S-TWA are less biased than their equivalent estimates produced with the SWA.

In addition, the two approaches of sub-group tailored weighting (*interaction-based and modelling response separately for each sub-group*) seem to be analogous in terms of their resultant weights. Their weights do not seem to affect estimates differently. Furthermore, the S-TWA has much more impact on estimates constructed from the sub-groups in question compared to estimates constructed from full sample analysis. In the latter case, although the S-TWA has resulted in a couple of considerable differences, our analysis here indicates that these changes may not be as significant as the ones from the analyses restricted to the sub-groups selected for the S-TWA. Still, these differences indicate that the S-TWA may affect total sample estimates, and with other sub-groups, it may be possible to prove that this effect can result in significant differences.

Table 3.10 Random effects logistic regression models of the determinants of the desire for residential mobility for those who were born in 1965 or after.

	Using SWs	**Using TWs_1**	**95% CI of (b_{SW} - b_{TW1})**		**Using TWs_2**	**95% CI of (b_{SW} - b_{TW2})**	
Years 1995 to 1998	1.137	1.175	-0.168	0.092	1.170	-0.163	0.097
Female	0.866*	0.868*	-0.220	0.126	0.864*	-0.216	0.220
White	1.041	1.081	-0.793	0.713	1.021	-0.320	0.360
Age	1.045	1.044	-0.003	0.005	1.044	-0.003	0.005
Member of a large household	1.057* [a]	1.133* [a]	-0.126	-0.026	1.136* [a]	-0.130	-0.028
Lives in a house with 3 to 4 rooms	1.242*	1.255*	-0.246	0.220	1.251*	-0.242	0.224
Lives in a house with 5+ rooms	1.365***	1.386***	-0.285	0.244	1.377***	-0.276	0.252
Has savings	0.979	0.971	-0.088	0.104	0.970	-0.087	0.087
House owned outright	0.845*	0.826*	-0.180	0.218	0.838*	-0.192	0.192
House owned with mortgage	1.562***	1.560***	-0.221	0.225	1.532***	-0.583	0.643
N	1,525	1,525			1,525		
σ	1.44	1.45			1.45		
ρ	0.38	0.39			0.39		

* The entries are odds ratios. All models are estimated by using a balanced panel of those who were born in 1965 or after and who responded in the first 8 waves. The reference categories of the independent variables are: years 1991 to 1994, male, non-white, member of a small household (3 members or less), lives in a house with 1 or 2 rooms, has no savings and tenant. [a] indicates a significant difference between the equivalent estimates adjusted with the *SWs* and both sets of *TWs*.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.11 Random effects logistic regression models of the determinants of the desire for residential mobility for the whole sample.

	Using SWs	**Using TWs_1**	**95% CI of (b_{SW} - b_{TW1})**		**Using TWs_2**	**95% CI of (b_{SW} - b_{TW2})**	
Years 1995 to 1998	1.165	1.164	-0.048	0.049	1.164	-0.049	0.051
Female	1.180*	1.171*	-0.120	0.138	1.181*	-0.130	0.128
White	1.241*	1.243*	-0.313	0.309	1.202*	-0.271	0.349
Age	0.964*	0.972*	-0.046	0.030	0.973*	-0.047	0.029
Member of a large household	1.053**	1.107**	-0.130	0.022	1.109**	-0.132	0.020
Lives in a house with 3 to 4 rooms	1.067*	1.068*	-0.132	0.130	1.083*	-0.147	0.115
Lives in a house with 5+ rooms	1.191**	1.195**	-0.153	0.145	1.207**	-0.165	0.133
Has savings	1.054	1.049	-0.053	0.063	1.53	-0.056	0.058
House owned outright	0.836***	0.854***	-0.314	0.278	0.869***	-0.325	0.259
House owned with mortgage	1.587***	1.569***	-0.125	0.161	1.579***	-0.135	0.135
N	6,753	6,753			6,753		
σ	2.25	2.27			2.26		
ρ	0.60	0.61			0.61		

* The entries are odds ratios. All models are estimated by using a balanced panel of those who responded in the first 8 waves. The reference categories of the independent variables are: years 1991 to 1994, male, non-white, member of a small household (3 members or less), lives in a house with 1 or 2 rooms, has no savings and tenant. * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$.

3.8 Conclusion

In this chapter we introduced an alternative approach (subgroup-tailored weighting) to create non-response weights in longitudinal studies. The subgroup-tailored weighting approach (S-TWA) is based upon selecting certain sub-groups from the survey sample and customising the construction of non-response weights to these sub-groups. Unlike the SWA, in the S-TWA, the weights are created by using a set of weighting variables that affects the response probability in the selected sub-groups regardless of whether or not it also affects the response probability in the rest of the sample. Also, the estimation of the weighting models in the S-TWA may be restricted to sample members from the sub-groups in question. Additionally, we introduced two possible approaches to carry out the S-TWA: *interaction-based approach and modelling the response propensity separately for each selected sub-group*.

The major findings of this chapter can be summarised in four main points:

1. The effect of the S-TWA on estimates is generally similar to that of the SWA, in particular in terms of estimates precision levels.
2. On some estimates, the S-TWA produces different results (in terms of magnitude) than the SWA.
3. It seems possible for the S-TWA to affect both total sample estimates and estimates derived only from the sub-groups selected for tailored weighting. However, the effect seems to be stronger (significant) on the estimates constructed from the sub-groups selected for the tailored weighting.

4. The two introduced approaches of S-TWA appear to produce similar sets of tailored weights that result in the same effect on estimates.

These findings encompass a number of propositions in the development of non-response weighting in longitudinal surveys. First, the findings suggest that the set of weights that can be produced from the S-TWA is somewhat different than the set of weights that results from the SWA. The difference emerged as a result of the different methodology followed to create the tailored weights. Changing the standard non-response covariates and restricting the weighting model to the sub-groups for which the tailored weights are created can result in a set of tailored weights that has different weight values than the standard weights. As a result, the tailored weights may drive some estimates to differ from their equivalent estimates constructed with standard weights. If both the changes in the non-response covariates and the set of respondents adopted in the S-TWA reflect the non-response process in the sub-groups in question better than the SWA, the S-TWA can be said to handle non-response in the sub-groups under investigation better than the SWA.

Second, although our investigation here does not show evidence that the S-TWA results in significantly different estimates than the SWA when estimates are derived from full sample analysis, it shows that some of the total sample estimates may still change considerably in terms of their magnitude if adjusted with the S-TWA. We believe that such changes, sometimes, have different impact on the interpretation of the results, especially with sensitive measures in some of the socio-economic processes. Hence, different conclusions regarding some of the total sample estimates could still be drawn on the basis of the S-TWA.

Third, our analysis suggests that the two approaches of the subgroup-tailored weighting (*interaction-based approach and modelling the response propensity separately for each selected sub-group*) may substitute one another. However, one may still expect differences – maybe not to a large extent - between these two approaches if they are applied on a different data set or different sub-groups. This is especially so if the number of the proposed variables for the tailored weighting is large. Thus, if the S-TWA is considered, we recommend the application of the second approach (*modelling the response propensity separately for each selected sub-group*) because it has some advantages over the first one (*interaction-based approach*). One of these advantages is that the second approach avoids the complications associated with too many interactions in the weighting model. Another advantage is that it allows restricting the weighting model to sample members in the sub-group selected for tailored weighting, which in turn permits excluding variables that do not predict response in the sub-group in question.

The availability of a large number of weighting variables in longitudinal surveys is advantageous. However, any weighting approach that depends on using a large number of variables to model the response propensity in the sample, but assumes that the effects of these variables are the same for different sub-groups (such as the SWA), may not always explain the non-response process well in all sub-groups in the sample. This is because samples in longitudinal surveys are large, and are often composed of units from a number of sub-populations which are not necessarily homogeneous in terms of the factors responsible for non-response. Successful weighting, in our opinion, depends on an independent and profound understanding of the non-response process in each of the major sub-groups in the sample rather than the number of variables included in a single weighting model. Even in the same survey sample, the cause of non-response may differ vastly across some sub-groups suggesting different sets of weighting variables (both in terms of scale and type) for weighting. Thus, looking at the non-response reasons in the sample as a whole may lead to ignoring variables that may appear insignificant in general while they are in fact important to explain non-response in some sub-groups. The findings in this chapter have demonstrated this. For example, it is known that factors like 'age' are powerful weighting variable while factors such as 'religion' are weak predictors of non-response; though, the results of this investigation showed the exact opposite within the subgroups on which we have focussed. At first glance, it may be hard to understand how a – well known - powerful auxiliary as 'age' could not be important in predicting response while a variable such as 'religion' is significant. However, once the cause of non-response is understood at a sub-group level, it can all be explained.

We expect similar findings if the S-TWA is applied in other panel studies, such as Understanding Society for example. However, in such a large longitudinal survey, the application of S-TWA might be, to some extent, tricky. This is because identifying the number of sub-groups that the tailored weighting should be based on is a subjective matter. S-TWA may be more appropriate for specific analyses where the analyst wants the best possible weights for a specific purpose. As for general-purpose public-release weights, it could be challenging to produce the best possible set of tailored weights because, in longitudinal survey samples, sub-groups maybe identified in a number of dimensions. Therefore, it would be difficult to identify a number of sub-groups that allows the execution of the best subgroup-tailored weighting. However, it should be pointed out that the more sub-groups used to create tailored weights (bearing in mind that the relevant sample sizes should be large enough to estimate non-response well) the stronger the effect of the overall set of tailored weights will be. Additionally, even if the number of the required sub-groups is accurately identified, the survey organisation will face the problem of identifying "which specific sub-groups should be used for tailored weighting?" as this maybe a subjective matter too.

Sub-groups can be non-overlapping (e.g. the sub-groups used in the analysis of this chapter). In this case, the sub-sets of tailored weights can be put together to form an Overall Set of Tailored Weights (OSTW). Accordingly, the OSTW can be beneficial in analyses that target the whole sample or analyses restricted to sub-groups. However, the sub-groups selected for tailored weighting maybe overlapping (e.g. sub-groups of males, disabled and white respondents). In this case, producing a OSTW is not possible via this method and, therefore, a number of sets of tailored weights may need to be released separately. However, this type of sub-group tailored weighting may not be appropriate for total sample estimates adjustment.

Thus, a future research may investigate a procedure that decides on:

- The number of sub-groups required for an effective overall set of tailored weights.
- Whether sub-groups should or should not be overlapped.
- Which specific sub-group should be selected for tailored weighting.

Finally, as the S-TWA uses a different set of variables compared to the one used in the SWA, researchers who are deciding between tailored weights and standard weights should pay attention to the set of variables used to create the tailored weights. This is because weights are also powerful in dealing with non-response bias if they are created using a set of variables that is strongly correlated with the main variable in the analysis (the dependent variable). Therefore, standard weights may also be a good choice if its weighting variables are more correlated with the dependent variable in the analysis. In this case it is a trade off between the reward of the tailored weights and the relationship between the dependent variable and the weighting variables used to create the standard weights. Thus, if a survey organisation considered S-TWA as an alternative, it may still want to keep standard weights in the public data files. Moreover, if a set of tailored weights is included in the data files, the survey organisation should properly document the process of weights creation as well as clearly stating the variables used to create the weights.

Conclusion

What have we studied?

The focus in this thesis has been devoted to non-response weighting in longitudinal surveys. It is typically the responsibility of survey organisations to design the weights and release them to be used by analysts. Thus, it is also the responsibility of survey organisations to ensure that the weights are created in the most appropriate manner by considering alternative approaches in the development of the weights. This study has been conducted to contribute to this area.

Most longitudinal surveys, nowadays, implement a similar approach in terms of non-response weighting for which this thesis has assigned the label 'the standard weighting approach' (SWA). The thesis has set out the principles of a typical SWA as: response is identified as responding in all waves up to the latest, and therefore weights are only provided for units responding in all waves up to the last one; non-respondents whose eligibility is unknown are assumed as eligible non-respondents, and are therefore included in the calculation of the weights; and only variables that distinguish response from non-response for the sample as a whole are used in the weights creation, and therefore variables that are only important at a subgroup level are ignored.

In return, one of the emphases in this thesis has been to disclose some weaknesses of the SWA with respect to its principles. In this regard, the thesis has raised three issues (corresponding to the three principles of the SWA) which are not taken into account by the SWA, and for which it has developed alternative weighting approaches (AWAs). These issues are: non-monotonic response pattern, unknown eligibility whilst weighting and subgroup-tailored weighting.

As has been demonstrated by this study, ignoring these issues in the process of producing the weights may yield less precise and, sometimes, biased estimates in substantive analyses that use the resultant weights. For survey organisations that apply the SWA, but are planning to embark on developing their non-response weighting, dealing with these issues should be part of that plan in order for weighting to more appropriately adjust the data and correct for non-response. In such cases, this thesis can be useful, as it suggests an alternative weighting approach (AWA) corresponding to each of the raised issues in conjunction with investigating the effects of ignoring these issues on weighting.

The study findings

By investigating the issues in question, a number of contributions to the SWA are made in this thesis. Some of the findings are chapter-specific and were discussed in details within the respective chapters (Weighting for Non-monotonic Response Pattern in Longitudinal Surveys, Unknown Eligibility whilst Weighting for Non-response: the Puzzle of who has Died and who is still Alive? and Non-response Subgroup-tailored Weighting: the Choice of Variables and the Set of Respondents Used to Estimate the Weighting Model). This section however will give a brief overview of the main finding in each chapter, and will synthesise these findings to answer the study's three questions.

Non-monotonic response patterns

The thesis showed that when a non-monotonic response pattern applies, the weights resulting from the SWA may contain zero weights for some of the responding sample members in samples drawn for analysis from a waves-combination that does not contain all waves. This is because the SWA designs the weights by defining response as responding at all waves up to and including the latest, whereas this is not necessarily the case for all sample members in the responding sample drawn from a possible combination of waves that does not contain all the waves. As a result, the SWA may isolate the influence of some responding sample members on estimates. It was therefore found that estimates resulting from the SWA are less precise, but they could also be biased if those respondents who are isolated by the SWA are different from those who are considered.

In contrast, the analysis showed that the AWA (introduced in the thesis) may result in more accurate estimates compared to the SWA. This is because the AWA designs the weights by defining response as responding to the waves-combination under investigation, and hence weights from this approach do not contain zero weights for sample members responding in this waves-combination. Corollary, there will be no responding cases of which their influence on the estimate in question will be isolated. In other words, the AWA does take into account the fact that the response pattern is non-monotonic. However, the evidences suggest that, overall, estimates resulting from the SWA and the AWA are similar and the differences are restricted to some estimates.

Unknown eligibility whilst weighting

It is unlikely that all sample members whose eligibility is unknown are eligible. However, the SWA does not recognise this. It assumes that all sample members for whom eligibility cannot be established during the estimation of the weights are eligible. Our investigation has shown that such treatment of unknown eligibility in the context of weighting may incorrectly result of larger weights as a consequence of including influence from ineligible sample members (who are assumed as eligible by the SWA) in the construction of the weighs. The results indicate that the larger weights could affect the magnitude of the estimates, but it may also introduce more variability in the set of weights resulting from the SWA. Consequently, some estimates resulting from the SWA could be biased or/and less accurate.

While it is not possible to identify eligibility status – at the case level - for all sample members, the thesis introduced another AWA by which it demonstrated how the effect of unknown eligibility (in terms of the larger weights) could be reduced. By reference to the population eligibility information from an external source, the AWA estimates the eligibility rate in the sample and use it to adjust the weights. The findings suggest that the adjusted weights result in similar estimates to the estimates produced from the SWA, but for some estimates, the AWA produced more precise and less biased estimates compared to the SWA.

Subgroup-tailored weighting

As it uses all sample members in weighting, the SWA relies on generic weighting variables that only predict the response propensity for the sample as a whole, and therefore it does not consider variables that are only important at a subgroup level. As non-response may differ (in terms of reasons) across subgroups in the same sample, the findings suggested that, for some subgroups, weights resulting from the SWA may not be the perfect adjustment for estimates based on these subgroups.

In this respect, the AWA tailors the weighting for some subgroups (by using specific variables and respondents in the subgroups in question to create the weights) in the sample. In general, the tailored-weights result in estimates that are similar to the estimates resulting from the SWA. However, for some estimates, the results indicate that the tailored-weights tackle non-response error – in the subgroups under consideration- better than weights based on the SWA (assuming that the variables used in the creation of the tailored-weights are more powerful in predicting response in the subgroups in question compared to the variables used in the SWA).

According to all the findings, the research questions in this study can be answered as follows:

1) Like any ordinary weighting approach, the SWA can deal adequately with the main aspects of the survey design. However, given that the SWA does not take into account a few important aspects of the survey that result from the longitudinal nature of the survey (non-monotonic response pattern, unknown eligibility and the choice of weighting variables and respondents), can the SWA deal with non-response error in all survey-based estimates?

The SWA cannot handle non-response error in all survey-based estimates. As a direct consequence of the complexity of longitudinal surveys coupled with the fact that the SWA does not take into account the three issues under investigation in this thesis, for some estimates, the SWA may not be successful in correcting the error.

2) If 'non-monotonic response pattern', 'unknown eligibility' and 'the choice of weighting variables and respondents' are taken into account to develop AWAs, will the AWAs have a different impact (in terms of magnitude and variance) on survey-based estimates compared to the SWA?

Yes. Because they are developed by taking these issues (issues mentioned in the question) into account, the AWAs can also tackle non-response error, in the estimates that are not adjusted appropriately by the SWA.

3) If the AWAs have a different impact on survey-based estimates as compared to the SWA, does this result in very different estimates (i.e. is the difference between the equivalent estimates resulting from the SWA and the AWAs significant)?

In general, the results arrived at through all of the AWAs are consistent with the results from the SWA. In other words, generally, it can be said that the effect of the AWAs on survey-based estimates is similar to that of the SWA indicating no odd outcome from the alternative approaches. However, the AWAs seem to affect some of the estimates in the analysis differently. These estimates appear to be adjusted more appropriately with the AWAs reflecting, either increase in the sample size in some cases (chapter 1), or changes in the weights' values and variance in other cases (chapter 2 and 3).

It is therefore the recommendation of this thesis that survey organisations consider the issues investigated here, and plan to implement the alternatives alongside the SWA for more development on non-response weighting.

Policy implication

Dealing appropriately with non-response requires paying attention to three aspects of the phenomena: the first is understanding the mechanism by which non-response occurs; the second is putting together an effective data collection protocol to reduce non-response (e.g. use of incentive and mixed-mode designs); and the third is designing an efficient approach to dealing with non-response bias in estimation, e.g. through weighting adjustment as suggested by the current study. However, despite the effort made to understand the mechanism of non-response (e.g. Goyder, 1987; Tourangeau *et al*, 2000), causes of non-response have not yet been understood at a "profound" level (Brick, 2013). Also, although data collection strategies that implement mixed-mode designs and incentives have been shown to increase response rates (e.g. De Leeuw, 2005; and Laurie and Lynn, 2009), a spate of research showed that the increase in response rates does not necessarily result in reduction in non-response bias (e.g. Curtin *et al*, 2000; and Groves, 2006). As a result, relyiance on post-survey adjustments, such as weighting, is increasing (Brick, 2013).

Therefore, survey organisations that apply weighting should pay more attention when establishing policies for weighting. One particular policy of the SWA is the production of a single set of weights that is based on a single modelling strategy. This policy aims at preparing one set of longitudinal weights at every wave and include it in the data file for public use. On the one hand, the aim of this policy is good since analysts who would like to use weights can easily find the weight-variable, and can use it in any analysis (multipurpose) since it is just one set (rather than digging in data files trying to find the appropriate weight-variable and might consequently end up with the wrong set of weights). On the other hand, the methodology of this policy may be of concern. Analysts always assume that the weight-variable reduces non-response error in any analysis and with any group of respondents from the sample. However, this is not necessarily the case. This study has used empirical findings to point out this fact.

The main problem of only implementing the SWA is centred around the fact that, even for the same survey, the approach to creating powerful and accurate weights may differ across different analyses objectives. Thus, the availability of alternatives such as the ones offered by this study is vastly useful. For sophisticated data analysts who understand the limitations of the SWA, but who may not have the data required for the construction of weights (e.g. initial probabilities of selection, auxiliary variables, etc…) that are most appropriate for their analyses, this is very convenient. For analysts who do not recognise the limitations of the SWA, the extra weights are a bonus. For survey organisations that design, collect and release longitudinal data, it is a development.

Thus, in longitudinal surveys, it is worthwhile to create a number of sets of weights, especially for wave-combinations that obtain data on the same subject. This suggestion supports the recommendations of Lynn and Kaminska (2010) when they proposed criteria for designing sub-sets of non-response weights. In their proposal, Lynn and Kaminska recommended that weights should be created for wave-combinations that are more likely to be used for analysis.

Also, using population information about survey eligibility to adjust the weights of the responding sample members is helpful. It reduces the weights of respondents who have similar characteristics to the ineligible sample members who were assumed to be eligible during the estimation of the weights. As a result, the influence of the ineligible sample members whose eligibility status is unknown will be reduced after the adjustment. This approach is similar – but not exactly the same - to that implemented in the Survey of Family, Income and Employment in New Zealand (SoFIE), which is presented by Statistics New Zealand (2011). SoFIE used information from New Zealand population census to determine the proportion of people who were not resident in the country (ineligible). The figures were used to develop benchmarks of counts by age and sex, and counts by ethnicity and age. Weights were then calibrated to benchmarks.

Additionally, although the idea of subgroup tailored-weighting is new and might have not been used before, the evidences from this study appear to support the fact that this approach is effective in reducing non-response error. This is because the tailored-weights can be designed using variables that are strongly related to the response probability in the sub-group under investigation. In the literature of weighting adjustment, there are a number of methods for choosing effective weighting variables when many variables are available. Brick (2013) provides a review of these methods. For example, search algorithm and regression models (Brick and Kalton, 1996). In these methods, the sample is divided into cells that discriminate between response and non-response or variables related with key outcome variables. These methods allow the identification of important variables interactions for bias reduction. Schouten (2007) is another example. He introduced a forward-backward strategy of variables selection similar to stepwise regression. These methods are improvement over the traditional approach of choosing weighting variables which relies on including demographic variables such as age, gender and geographical area even if they are not effective in reducing bias (Peytcheva and Groves, 2009). However, they still do not take into account the fact that the relation between a set of weighting variables and the response propensity may change dramatically across sub-groups in the sample. Therefore, combining any of these methods with the approach introduced in this study could be advantageous for selecting effective set of weighting variables.

In short, from the analysis point of view, a weighting policy that rests on implementing alternative approaches beside the SWA may be desirable and, above all, it could be very effective in dealing with non-response. From the survey organisation's point of view, this may require re-establishing an existing weighting policy. It means extra time and effort to estimate more weighting models, involve population information to calibrate the weights and proper documentation on the new weighting before including weights in the data files. However, it is also a reflection of the development in the survey organisation in terms of the quality of what is offered to the public.

Future research

The complexity of longitudinal surveys makes the improvement of non-response weighting challenging. The scale of investigating alternatives should be extensive and multidimensional. To promote the understanding and generate comprehensive strategies with regard to weighting, there is a need for more research to allow further assessments. Exploring the following as future research directions can facilitate the attainment of this goal:

- Reapplying the same approaches in this study on similar longitudinal data sets (e.g. data from the German Socio-economic Panel) may strongly support the findings of this study or otherwise lay out the foundation for different considerations.
- Replicating the approach in chapter 1 on a different combination of waves (with different questionnaire topics); repeating the analysis in chapter 2 with different

substantive analyses; and reapplying the weighting strategy in chapter 3 on different sub-groups in the sample

- The current study has introduced a number of alternative approaches; a future research may investigate a possibility of combining these techniques into one standardised and comprehensive weighting approach that can guide the development of non-response weights.

Final statement

Despite the benefit of adjustment weighting in terms of reducing non-response bias, in practice, and particularly in longitudinal surveys, weighting has encountered some obstacles (such as the ones investigated here) that precluded some of its advantages. In complex longitudinal surveys, weighting might not achieve all of the anticipated if it was done with an ordinary approach such as the SWA. Although the SWA can be helpful for a number of substantive analyses, in some cases its benefits may not be comprehensive as has been shown by this study. So, is it wise for new surveys to consider multiple approaches while establishing their weighting schemes? Is it worthwhile for existing surveys that only apply the SWA to introduce additional weighting approaches after all the waves that have been conducted? Would data analysts, who use weights in their analysis, prefer to continue using weights from the SWA or would they rather have the opportunity to alternate with weights based on the AWAs? And, if developing alternative weighting was intended in some survey organisations, should our introduced AWAs be considered in this development?

References

AAPOR: The American Association for Public Opinion Research (1998) "Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for RDD Telephone Surveys and In-person Household Surveys". Ann Arbor, Michigan: AAPOR.

AAPOR: The American Association for Public Opinion Research (2004) "Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys", 3rd Edition. Lenexa, Kansas: AAPOR,[online] available fromwww.aapor.org/pdfs/standarddefs2004.pdf [Accessed 5th January 2015].

Allison, P. (2000) "Missing Data". Thousand Oaks, CA: Sage.

Alreck, P., and Settle, R. (1995) "The Survey Research Handbook", New York: McGraw-Hill.

Alsnih, R., and Stopher, P. (2004) "Estimating Eligibility Rates: A crucial Component of the Calculation for Response Rates", *Institute for Transport Studies Working Paper*, NO ITS-WP-04-08. Sydney: University of Sydney.

Atrostic, B.K, Bates, N., Burt, G., and Silberstein, A. (2001) "Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends and New Insights", *Journal of Official Statistics*, Vol. 17, pp. 209-226.

Bates, N., Dahlhamer, J., and Singer, E. (2008) "Privacy Concerns, Too Busy, or Just Not Interested: Using Doorstep Concerns to Predict Survey Nonresponse", *Journal of Official Statistics*, Vol. 24, pp. 591-612.

Becketti, S., Gould, W., Lillard, L., and Welch, F. (1988) "The Panel Study of Income Dynamics after fourteen years: an evaluation", *Journal of Labour Economics*, Vol. 6, pp. 472-492.

Beerten, R., Lynn, P., Laiho, J., and Martin, J. (2000) "Response Rates as a Measure of Survey Quality", [online] available fromhttp://webfarm.jrc.cec.eu.int/ETK-NTTS/Papers/final-papers/61pdf. [Accessed 5th January 2015].

Bethlehem, J.(2009) "Applied Survey Methods: A Statistical Perspective*",* New Jersey: John Wiley& Sons, Inc.

Biemer, P. P., and Christ, S. L. (2008) "Weighting Survey Data", (chapter 17) in De Leeuw, E.D., Hox, J.J., and Dillman, D.A., *International Handbook of Survey Methodology*. New York: Lawrence Erlbaum.

Brandon, E. (2012) "Why Older Citizens are More Likely to Vote".[online] available from http://money.usnews.com/money/retirement/articles/2012/03/19/why-older-citizens-are-more-likely-to-vote [Accessed 23d June 2013].

Brehm, J. (1993)"The phantom respondents: Opinion surveys and political representation". Ann Arbor: University of Michigan Press.

Brick, J. (2013) "Nonresponse and Weighting Adjustments: A critical review", *Journal of Official Statistics*, Vol. 29(3), pp. 329-353.

Brick, J., and Kalton, G. (1996) "Handling Missing Data in Survey Research", *Statistical Methods in Medical Research*, Vol. 5, pp. 215- 238.

Carpenter, J., and Plewis, I. (2011) "Analysing Longitudinal Studies with Non-response: Issues and Statistical Methods", (pp. 518-543) in Williams, M., and Vogt, W., *Hand Book of Methodological Innovation*. Newbury Park: Sage.

CASRO: Council of American Survey Research Organizations (1982), "On the

Definition of Response Rates: A Special Report of the CASRO Task Force on Completion Rates", [online] available fromwww.casro.org[Accessed 25th December 2014].

Chang, M. (2010) "Adjusting for Nonresponse Bias Due to Attrition and Intermittent Wave Nonresponse in Longitudinal Surveys of Older Adults", *Dissertation*. Nebraska: University of Nebraska.

Cochran, W. (1977) "Sampling Techniques", John Wiley and Sons.

Collins, L. M., Schafer, J. L., and Kam, C. M. (2001)"A comparison of inclusive and restrictive strategies in modern missing data procedures", *Psychological Methods* , Vol. 6, pp. 330-351.

Couper, M. P. (1997)"Survey introductions and data quality", *Public Opinion Quarterly*,Vol. 6, pp. 317-338.

Curtin, R., Stanley, P., and Singer. E. (2000) "The Effects of Response Rates Changes on the Index of Consumer Sentiment", *Public Opinion Quarterly*, Vol. 64, pp413-28.

David, M., Little, R. J. A., Samuhel, M. E., and Triest, R. K. (1983) "Non-random Non-response Models Based on the Propensity to Respond", *Proceedings of the Business and Economic Statistics Section*, *American Statistical Association*, pp. 168-173.

Deane, G. (1990). "Mobility and adjustments: paths to the resolution of residential stress", *Demography*, 27, pp. 65–79.

De Leeuw, E., and De Heer, W. (2002) "Trends in Household Survey Nonresponse: A longitudinal and International Comparison", in Groves, M., Dillman, A., Eltinge, L., and Little, A. New York: Wiley, pp. 41-54.

De Leeuw, E., Hox, J., and Huisman, M. (2003) "Prevention and Treatment of Item Non-response", *Journal of Official Statistics,* Vol. 19(2), pp. 153-176.

De Leeuw, E. (2005) "To Mix or Not to Mix Data Collection Modes in Surveys", *Journal of Official Statistics*, Vol. 21(2), pp. 233-255.

Deville, J. C., and Särandal. C. E. (1992) "Calibration Estimators in Survey Sampling*",* *Journal of the American Statistics Association*, Vol. 87, pp. 376-382.

Dillman, D. (2007) "Mail and internet surveys: The tailored design method*",* New York: John Wiley & Sons Inc.

Dr Foster. (2004) "Death Rates in England and Wales and the United States: variation with age, sex and race", *BMJ*, vol.329, pp. 1367-1367.

Duncan, G., Hofferth, S. L., and Stafford, F. P. (2004) "Evolution and Change in Family Income, Wealth and Health: The Panel Study of Income Dynamics, 1968-2000 and beyond". (chapter 6) In House, J. (Eds), *A telescope on Society*. Ann Arbor: Institute for Social Research, University of Michigan.

Ellis, J. (2000) "Estimating the Number of Eligible Respondents for a Telephone Survey of Low-Incidence Households". [online] available fromhttp://www.amstat.org/sections/srms/proceedings/papers/2000_180.pdf[Accessed 12th June 2011].

Fitzgerald, J., Gottschalk, P., and Moffitt, R. (1998)"An Analysis of Sample Attrition in Panel Data: the Michigan Panel Study of Income Dynamics", *Journal of Human Resources*, pp.251-299.

Frankel, R. (1983) "The Report of the CASRO Task Force on Response Rates", (pp. 1-11) in Wiseman, F., *Improving Data Quality on Sample Surveys*. Cambridge: Marketing Science Institute.

Frankel, R., Battaglia, P., Kulp, W., Hoaglin, C., Khare, M., and Cardoni, J. (2003) "The Impact of Ring-No-Answer Telephone Numbers on Response Rates in Random-Digit Dialing Surveys", Presentation to the American Statistical Association, San Francisco.

Gasteiz, V. (2007) "Standardisation and Systematisation of Response Rate Calculation", *Euskal Estatistika Erakuundea, Basque Statistics Office*. [online] available fromhttp://www.eustat.es/document/datos/ct_16_i.pdf[Accessed 7th May 2011].

Cervantes, I., and Brick, M. (2008) "Empirical Evaluation of Raking Ratio Adjustments for Non-response", *Joint Statistical Meeting*, Denver, Colorado: Section on Survey Research Methods*"*.[online] available fromhttps://www.amstat.org/sections/SRMS/Proceedings/y2008/Files/301418.pdf[Accessed 25th December 2014].

Glenn, N., and Frisbie, W. (1977) "Trend Studies with Survey Sample and Census Data", *Annual Review of Sociology*, Vol. 3, pp. 79-104.

Goldstein, H. (2009) "Handling Attrition and Non-response in Longitudinal Data", *Longitudinal and Life Course Studies*, Vol. 1, pp. 63-72.

Gouskova, E. (2001) “PSID Technical Report: The 2002 PSID Child Development Supplement (CDS-II) Weights*”*.[online] available fromhttp://psidonline.isr.umich.edu/data/weights/[Accessed 19th December 2010].

Goyder, J. (1985) “Face-to-Face Interviews and Mail Questionnaires: The Net Difference in Response Rate”, *Public Opinion Quarterly*, Vol. 49, pp. 234-252.

Goyder, J. (1987) “The Silent Minority: Nonrespondents on Sample Surveys”, Boulder, CO: West View Press.

Gray, R., Campanelli, P., Deepchand, K., and Prescott-Clarke, P. (1996), “Exploring Survey Non-response: the effect of attrition on follow-up of the 1984-85 health and life style survey”, *The Statistician*, vol. 45, pp. 163-183.

Grau, E., Potter, F., Williams, S., and Diaz-Tena, N. (2006) “Nonresponse Adjustment Using Logistic Regression: to weight or not to weight”, New Jersey: Mathematica Policy Research, Inc.[online] available fromhttp://www.amstat.org/sections/srms/proceedings/y2006/Files/JSM2006-000717.pdf[Accessed 19th October 2010].

Gray, R., Campanelli, P., Deepchand, K., and Prescott-Clarke, P. (1996) “Exploring Survey Non-response: the effect of attrition on follow-up of the 1984-85 Health and Life Style Survey”, *The Statistician*, Vol. 45, pp. 163-183.

Groves, R. M. (1989)“Survey errors and survey costs”.New York: John Wiley & Son.

Groves, R., and Couper, M. (1998) “Nonresponse in Household Interview Surveys”, New York: Wiley.

Groves, R., Presser, S., and Dipko, S. (2004) "The Role of Topic Interest In Survey Participation Decisions", *Public Opinion Quarterly*, Vol. 68, pp. 2-31.

Groves, R., Fowler,F., Couper, M., Lepkowski, J., Singer, E., and Tourangeau, R.(2004)"Survey Methodology", New York: Wiley.

Groves, R. (2006) "Nonresponse Rates and Nonresponse Bias in Household Surveys", *Public Opinion Quarterly*, Vol. 70, pp. 646-675.

Groves, R., Wagner, J., and Peytcheva, E. (2007) "Use of Interviewer Judgments About Attributes of Selected Respondents in Post-survey Adjustment for Unit Nonresponse: An Illustration With the National Survey of Family Growth", *Proceedings of the Survey Research Methods Section, ASA.*

Groves, R., and Peytcheva, E. (2008) "The Impact of Nonresponse Rates on Nonresponse Bias: A meta Analysis", *Public Opinion Quarterly*, Vol. 72 (2), pp. 89-167.

GTSS(Global Tobacco Surveillance System). (2009) "Global Adult Tobacco Survey (GATS): sample weights manual". [online] available from www.cdc.gov/tobacco/global/gats. [Accessed 5th October 2010].

Hausman, J. A., and Wise, D. A. (1979)"Attrition bias in experimental and panel data: The Gary income maintenance experiment", *Econometrika* , Vol.47, pp. 455-473.

Hawkes, D., and Plewis, I. (2006) "Modelling Non-response in the National Child Development Study", Journal of the Royal Statistical Society Series A, Vol. 169 (3), pp. 479-491.

Heckman, J. (1979) "Sample Selection Bias as a Specification Error". *Econometrica*, Vol. 47, pp. 153-162.

Horvitz, D. G., and Tompson, D. J. (1952) "A generalisation of Sampling without Replacement from a Finite Universe" *Journal of the American Statistical Association*, Vol (47), pp. 663-685.

Johnson, W., and Krueger, R. 2006 "How Money Buys Happiness: Genetic and Environmental Processes Linking Finances and Life Satisfaction". *Journal of Personality and Social Psychology*, vol 90, pp. 680–691.

Jones, A., Koolman, X., and Rice, N. (2004) "Health-related Attrition in the BHPS and ECHP: using inverse probability weighted estimators in nonlinear models", [online] available fromhttp://www.ccsr.ac.uk/methods/events/attrition/documents/Jones.pdf[Accessed 17th July 2013].

Kalton, G., and I. Flores-Cervantes. (2003). "Weighting Methods.", *Journal of Official Statistics*, Vol. 19, pp. 81-97.

Kan, M., and Laurie, H. (2010) "Savings, Investments, Debts and Psychological Well-being in Married and Cohabiting Couples", *ISER Working Papers Series*, No. 2010-42. Colchester: University of Essex.

Kish, L. (1965) "Survey Sampling", New York: John Wiley and Sons.

Kim, J., and Moen, P. (2002) "Retirement Transitions, Gender, and Psychological Well-being: A life-course, Ecological Model", Journal of Gerontology, Vol. 57B(3), pp. 212-222.

Kohler, H., Behrman, J., and Skytthe, A. (2005) "Partner + Children = Happiness? The Effects of Partnerships and Fertility on Well-Being" *Population and development review*, Vol. 31, issue 3, pp.407-445.

Kreuter, F., Lemay, M., and Casas-Cordero, C. (2007) "Using Proxy Measures ofSurvey Outcomes in Post-survey Adjustments: Examples From the European Social Survey (ESS)", *Proceedings of the Survey Research Methods Section, ASA.*

Kreuter, F., and Krieger, U. (2009) "Analysing Contact Sequences in Call Record Data. Potential and Limitations of Sequence Indicators for Nonresponse Adjustments in the European Social Survey", *Journal of Official Statistics,* Vol. 25(2), pp. 203-226.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R., and Raghunathan, T.(2010) "Using Proxy Measuresand Other Correlates of Survey Outcomes to Adjust for Nonresponse: Examples From Multiple Surveys", *Journal of the Royal Statistical Society*, SeriesA, Vol. 173, pp.389-407.

Kreuter, F., and Olson, K. (2011) "Multiple Auxiliary Variables in Non-response Adjustment", *Sociological Methods & Research*, Vol.40, pp. 311–332.

Kroh, M. (2009) "Data Documentation: sample sizes and panel attrition in the German Socio Economic Panel (1984 until 2008)", Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).

Laurie, H., Smith, R., and Scott, L. (1999) "Strategies for Reducing Nonresponse in a Longitudinal Panel Survey" *Journal of Official Statistics*, vol.15, pp. 269-282.

Laurie, H., and Lynn, P. (2009) "The Use of Respondent Incentives on Longitudinal Surveys", in Lynn, P., *Methodology of Longitudinal Surveys.* Sussex: John Wiley & Sons, Ltd.

Lepkowski, J., and Couper, M. (2002) "Nonresponse in the Second Wave of Longitudinal Household Surveys", in Groves, R., Dillman, D., Eltinge, J., and Little, R., *Survey Nonresponse*. New York: John Wiley.

Lessler, J., and Kalsbeek, W. (1992) "Nonsampling Error in Surveys", New York, John Wiley & Sons, Inc.

Little, R. (1986) "Survey Nonresponse Adjustments for Estimates of Means." *International Statistical Review*, Vol. 54, pp.139-57.

Little, R. J. A., and Rubin, D. B. (2002) "Statistical Analysis with Missing Data". New York: Wiley.

Little, R. J. A., and Rubin, D. B. (2002) "Statistical Analysis with Missing Data". New York: Wiley.

Little, R. J., and Vartivarian, S. (2003) "On Weighting the Rates in Non-response Weights", *Statistics in Medicine*, Vol. 22, pp.1589-599.

Little, R. J., and Vartivarian, S. (2005) "Does Weighting for Nonresponse Increase the Variance of Survey Means?",*Survey Methodology*, Vol.31, pp.161-68.

Lynn, P. (1996) "Weighting for Non-response", pp. 205-214 in Banks. R. *et* al. *Survey and statistical computing 1996*, ASC.

Lynn, P. (2003) "PEDAKSI: Methodology for Collecting Data About Survey Nonrespondents", *Quality and Quantity*, Vol. 37, pp.239-261.

Lynn, P. (2005) "Weighting", pp.967-973 in Kempf-Leonard. K. *Encyclopedia of Social Measurment,* Academic press.

Lynn, P. (ed.). (2006) "Quality Profile: British Household Panel Survey. Version 2.0: Waves 1to 13: 1991-2003". Colchester: Institute for Social and Economic Research, University of Essex. [online] available from http://www.iser.essex.ac.uk/ulsc/bhps/quality-profiles/BHPS-QP-01-03-06-v2.pdf[Accessed 4th November 2010].

Lynn, P. (2008) "The Problem of Nonresponse",(pp. 35-55) in De Leeuw, E.D., Hox, J.J., and Dillman, D.A., *International Handbook of Survey Methodology.* New York: Lawrence Erlbaum.

Lynn, P. (2009) "Methods for Longitudinal Surveys", (chapter 1) in Lynn, P., *Methodology of Longitudinal Surveys.* Sussex: John Wiley & Sons, Ltd.

Lynn, P., and Kaminska, O. (2010) "Criteria for Developing Non-response Weight Adjustments for Secondary Users of Complex Longitudinal Surveys",*2010 International Workshop on Household Survey Non-Response.* Nürnberg.

Lynn, P. (2011) "Maintaining Cross-Sectional Representativeness in a Longitudinal General Population Survey", *Institute for Social and Economic Research Working Papers*, No. 2011-04. Colchester: University of Essex.

Lynn, P., Kaminska, O., and Goldstein, H. (2014) "Panel Attrition: How Important is it to Keep the Same Interviewer?", *Journal of official statistics*, Vol. 30, pp.443-457.

Madow, W., Nisselson, H., and Olkin, I. (1983) "Incomplete Data in Sample Surveys", New York: Academic Press Inc.

Massey, J. T. (1988) "An overview of telephone coverage", in Groves, R., Biemer, P., Lyberg, L. E., Massey, J. T., Nicholls, W. L., and Waksberg. J, *Telephone Survey Methodology*. New York: John Wiley & Sons, Inc.

McGonagle, K., Couper, M., and Schoeni, R. (2011) "Keeping Track of Panel Members: an experimental test of a between-wave contact strategy", *Journal of Official Statistics*, vol.27, No. 2, pp. 319-338.

McHugh, K. E., Gober, P., & Reid, N. (1990) "Determinants of Short- and Long- Term Mobility Expectations for Home Owners and Renters", *Demography*, 27(1), pp. 81-95.

Mohadjer, L., and Choudhry, G. (2001) "Adjusting for Missing Data in Low-Income Surveys", in Ploeg, M., Moffitt, R., and Citro, C. *Studies of Welfare Populations: Data Collection and Research Issues.* Washington: The National Academies Press.

Merkle, D., and Murray, E. (2002) "Nonresponse in Exit Polls: A comprehensive Analysis", in Groves, R., Dillman, D. A., Eltinge, J., and Little, R. J. A., *Survey Nonresponse*. New York: Wiley.

National Statistics. (1999) "Mortality Statistics: Review of the Registrar General on Deaths in England and Wales, 1999", *A national Statistics Publication Series DH1*, NO.32. London: Her Majesty's Office (HMSO).

Nicoletti, C., and Buck, N. (2004) “Explaining Interviewee Contact and Co-operation in the British and German Household Panels”, (pp. 143-166) in Ehling, M., and Rendtel, U. *Harmonisation of Panel Surveys and Data Quality*. Wiesbaden: Statistisches Bundesamt.

Nicoletti, C., and Peracchi, F. (2005) “Survey Response and Survey Characteristics: microlevel evidence from the European Community Household Panel”, *Journal of the Royal Statistical Society*, Series A, Vol. 168, pp.763-781.

Peytcheva, E. and Groves, M. (2009) “Using Variation in Response Rates of Demographic subgroups as Evidence of Nonresponse Bias in Survey Estimates”, *Journal of Official Statistics*, Vol. 25, pp. 193-201.

Plaza, S., and Graf, E. (2008) “Executive Summary: Recommendations and Practical Examples for Using Weighting”, Lausanne: University of Lausanne.

Plewis, I. (2011) “Auxiliary Variables and Adjustments for Missingness in Longitudinal Studies”, *the Cathie March Centre for Census and Survey Research*, eScholar ID: 138810. Manchester: University of Manchester.

Rabe, B., & Taylor, M. (2009) “Residential Mobility, Neighbourhood quality and life-course events”, *Institute for Social and Economic Research Working Papers*, No. 2009-28. Colchester: University of Essex.

Ryan,R., and Frederick,C. (2006) “On Energy, Personality, and Health: subjective vitality as a dynamic reflection of well-being”, *Journal of personality*, Vol. 65, issue 3, pp.529-565.

Rogers, G., Hummer, A., and Nam, B. (2000) "Living and dying in the USA. Behavioral, health, and social differences in adult mortality', New York: Academic Press.

Rosenbaum, P. R., and Rubin, D. B. (1983) "The Central Role of the Propensity Score in Observational Studies for Casual Effects", Biometrika, Vol. 70, pp. 41-55.

Rubin, D. B. (1976) "Inference and Missing Data (with discussion). Biometrika, Vol. 63, pp. 581-592".

Sadig, H., and Banany, A. (2015) "The Impact of Non-response Weighting: Empirical Evidence From Modelling Residential Mobility", *Journal of Social Research and Policy,* (Forthcoming).

Sanbonmatsu, L., Ludwig, J., Katz, L. F., Gennetian, L. A., Duncan, G. J., Kessler, R. C. (2011) "Moving to Opportunity for Fair Housing Demonstration Program - Final Impacts Evaluation". Cambridge, MA: National Bureau of Economic Research.

Särndal, C. E., and Lundström, S. (2005) "Estimation in Surveys with Nonresponse", Chichester: John Wiley & Sons Ltd.

Scheaffer, R., Mendenhall, W., and Ott, R. (1996) "Elementary Survey Sampling", California: Wadsworth.

Schouten, B. (2004) "Adjustment for Bias in the Integrated Survey on Household Living Conditions (POLS) 1998", *Statistics Netherlands Discussion Paper*, NO. 04001. Voorburg: Statistics Netherlands.

Schouten, B. (2007) “A selection Strategy for Weighting variables Under a Not-missing-at-random Assumption”, *Journal of Official Statistics*, Vol. 23, pp. 51-68.

Siddiqui, O., Flay, B., and Hu, F. (1996)“Factors Affecting Attrition in a Longitudinal Smoking Prevention Study”, *Preventive Medicine,*Vol. 25**,**pp. 554-560.

Sikkel, D., Hox, J., and De Leeuw, E. (2009) “Using Auxiliary Data for Adjustment in Longitudinal Research”. (chapter 9) in Lynn, P., *Methodology of Longitudinal Surveys.* Sussex: John Wiley & Sons, Ltd.

Singh-Manoux, A., Gueguen, A., Ferrie, J., Shipley, M., Martikainen, P., Bonenfant, S., Goldberg, M., and Marmot, M. (2008) “Gender Differences in the Association Between Morbidity and Mortality Among Middle-Aged Men and Women”, *American Journal of Public Health*, vol. 98, No. 12, pp. 2251-2257.

Smith, T. W. (1984)“Estimating nonresponse bias with temporary refusals”, *Sociological Perspectives* , Vol. 27, pp. 473-489.

Smith, T. (2003) “Areview of Methods to Estimate the Status of Cases with Unknown Eligibility”, *Report Presented to AAPOR, Phoenix, May, 2004.*

Speare, A. (1974) “Residential Satisfaction as an Intervening Variable in Residential Mobility”, *Demography*, 11(2), pp. 173-188.

Statistics New Zealand. (2011) “New Weighting Methodology in Longitudinal Surveys: as applied in the Survey of Family, Income and Employment”, Wellington: Statistics New Zealand. [online] available fromhttp://www.stats.govt.nz/browse_for_stats/income-and-work/Income/sofie/new-weighting-methodology-for-sofie.aspx, [Accessed 13th October 2013].

Steeth, C. (1981) “Trends in Non-response Rates 1952-1979”, *Public Opinion Quarterly*, Vol. 45, pp. 40-57.

Stoop, L. (2005) “The Hunt for the Last Respondent”, *The Hague, Netherlands: Social and Cultural Planning Office.*

Stoop, I., Billiet, J., Koch, A., and Fitzgerald, R. (2010) “Improving Survey Response: Lessons Learned from the European Social Survey”, Chichester: John Wiley & Sons, Ltd.

Summerfield, M. (ed.). (2010) “HILDA User Manual- Release 9”, Melbourne Institute of Applied Economics and Social Research, University of Melbourne.

Taylor, M. (ed.). (2010) “British Household Panel Survey User Manual Volume A”. Colchester: Institute for Social and Economic Research, University of Essex.

Tourangeau, R., Rips, J., and Rasinski, K. (2000) “The Psychology of Survey Response”, New York: Cambridge University Press.

Uhrig, N. (2008)“The Nature and Causes of Attrition in the British Household Panel Survey”, *Institute for Social and Economic Research Working Papers*, No. 2008-05. Colchester: University of Essex.

Watson, D., and Fry, T. (2002) “The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Wave 1 Weighting” *HILDA Project Technical Paper Series,* NO. 3/02. Melbourne: Melbourne Institute of Applied Economics and Social Research, University of Melbourne.

Watson, D. (2003) “Sample Attrition between Waves 1 and 5 in the European Community Household Panel”, *European Sociological Review*, vol. 19:4, pp.361-378.

Watson, D. (2004) “Wave 2 Weighting”, *HILDA Project Technical Paper Series*, NO. 4/04. Melbourne: Melbourne Institute of Applied Economics and Social Research, University of Melbourne.

Watson, N., and Wooden, M. (2004) “The HILDA Survey Four Years on”, *Australian Economic Review*, Vol. 37, pp.343-349.

Watson, N., and Wooden, M. (2009) “Identifying Factors Affecting Longitudinal Survey Response”, (chapter 10) in Lynn, P., *Methodology of Longitudinal Surveys.* Sussex: John Wiley & Sons, Ltd.

Watson, D. (2014) “Outcomes from Matching the HILDA Survey Sample to the Death Register” *HILDA Project Technical Paper Series*, NO. 2/14. Melbourne: Melbourne Institute of Applied Economics and Social Research, University of Melbourne.

Appendices

Chapter 1 appendices

Appendix A.1: Earlier version of chapter 1 which was presented in the Second Italian Conference on Survey Methodology (ITACOSM), 2011.

Non-response Weight Adjustments in Longitudinal Surveys

Husam Sadig[21]

Keywords: Non-response Error, Bias, Precision, Weighting.

1. Introduction

The multi-wave feature in longitudinal surveys allows for data to be drawn for analysis from different combinations of waves. However, the set of responding units can differ across wave-combinations offering potentially different subsamples for every possible combination. Thus, weights may be required for a number of combinations of waves too, as one set of weights might not be sufficient in handling non-response error in all subsets of data.

However, in the major longitudinal surveys in the world weighting for non-response is a single weighting strategy overlooking the fact that different wave-combinations can potentially provide different sets of respondents. In a single weighting strategy, weights are designed based on respondents from all waves up to the latest but used for analysis with data from any combination of waves. For instance, in the British Household Panel Survey (BHPS), longitudinal weights at any wave 'w' are only available for a balanced panel from all waves up to wave 'w' (Taylor *et al*, 2010). Likewise, longitudinal weights in a current wave in the Swiss Household Panel (SHP) are designed to extrapolate to the population living in Switzerland at that wave using respondents from all waves up to the current (Plaza and Graf, 2008). This is also the case in the German Socio Economic Panel (GSOEP) and the Panel Study of Income Dynamics (PSID), where no particular combination of waves are provided with specially designed longitudinal weights; instead, weights in the latest wave are available for the set of respondents from all waves including the latest (Kroh, 2009; Gouskova, 2001).

This single non-response weighting strategy, which is used in almost every survey, could be helpful and practical in reducing non-response bias, but may be inadequate in respect to the subsample being used for analysis. For example, for analytical purposes, data can be obtained for analysis from all waves or only subset of waves. Potentially, the set of responding participants is different in each case. Thus, a set of weights based on respondents at all waves will surely ignore any

[21]Institute for social and Economic Research (ISER), University of Essex, Colchester, UK, CO4 3SQ, hesadi@essex.ac.uk

group of respondents who are not present in all waves. Therefore, it is suboptimal if used with data from a subset of waves. The sub-optimality is based on the loss of respondents who should contribute to the estimation process. Furthermore, if those who responded to a particular combination of waves are systematically different (in terms of what is being measured) from those who responded at all waves, non-response bias might not be removed despite the use of non-response weights.

In theory, the way out of this problem is to design a subset of non-response weights for every possible combination of waves. However, providing weights for all possible combinations of waves might not be achievable in practice sometimes. For example, after k waves are conducted, there is a $(2^k$-1) possible combination of waves to provide weights for. Moreover, this number increases rapidly when more waves are added, and it could even outnumber the number of variables in the survey in a long term panel. However, in practice, not every possible combination of waves is of use for researchers. Therefore, only desirable subsets of weights should be produced. Nevertheless, it is a challenging task to identify combinations of waves that will be of interest for data users. But the possibility that a single weighting strategy might not be sufficient generates interest in the development of more subsets of weights. Hence, the investigation of this is an important aspect of weighting panel data.

Very little work has been done in this area. In fact, the only effort I have come across is by Lynn and Kaminska (2010), suggesting criteria for developing subsets of longitudinal non-response weights.

A common feature of longitudinal surveys is a frequently asked module of questions where certain waves are conducted to obtain information about specific topic(s). For example, wave 8, 13 and 18 in the BHPS provide data on neighborhood, expectations of relationships and marriage in future. Thus, it might be useful to provide BHPS data users with a subset of weights designed specifically for the analysis of data from these waves.

In this paper, I use data from wave 1 to 15 from the British Household Panel Survey (BHPS) to investigate whether the use of a single weight adjustment in longitudinal surveys is adequate to handle non-response error. I also evaluate the choice of providing a subset of weights to a combination of waves that carry the same module of questions.

2. METHODOLOGY

This paper used data from wave 5, 10 and 15 (waves collect data about wealth, assets and debts) from the BHPS and designed a subset of non-response weights for this combination. Also, another set of non-response weights is designed based on respondents at all waves up to wave 15. Analysis was carried out on savings and debts data from wave 5, 10 and 15 using the two sets of weights. The issue of interest here is to compare estimation results produced from the use of the two sets of weights.

Both sets of weights were created using a model based method. The analysis was restricted to respondents aged 16 or above and alive during the course of the 15 waves. A large mixture of continuous and categorical variables from wave 1 was used to model the response propensity in the two wave-combinations and create the weights. These variables were chosen from three categories of variables that are thought to affect the response propensity. These are: interview and interviewer characteristics (e.g. interviewer's sex and length of interview), household characteristics (e.g. household size and household type) and individual characteristics (e.g. age, sex and savings) from wave 1 were used to estimate each model.

The British Household Panel Survey provides detailed information on savings and debts at the individual level for the years 1995, 2000 and 2005, representing waves 5, 10 and 15 respectively. In each of these waves, respondents were asked if they have money in savings and whether they owe money. Based on this setting, tow random effects logistic regression models were used to estimate the determinants of having money in savings or being in debts respectively. However, each model was estimated twice using the two different longitudinal sets of weights. The main idea is to assess the change on the regression coefficients when varying weight adjustments procedures. In particular, the point of interest is to spot the influence of creating non-response longitudinal weights based on the consideration of combination of waves with the same module of questions.

3. RESULTS

As seen in Table 1., there is much to be learnt from the comparison between models. For instance, the sample size associated with the use of weights based on the respondents from all waves (4,654) is smaller than the sample size associated with the weights based on respondents from wave 5, 10 and 15 (5,132) by 478 respondents. This is because the former set of weights assigns a weight of zero to any case that is not present in all the 15 waves.

Focussing on models concerned with savings, having a second job and being unemployed are significant in the first but not the second model. This is clearly showing the effect of the increase in the sample size used to estimate the first model on these particular variables. In other words, using a weights adjustment method based on respondents in all waves which is associated with the loss of 478 respondents in the sample, results in underestimating the importance of having a second job and being unemployed. Moreover, although living with a partner is not significant in any of the two models, the signs of the coefficients in the two models are different.

As for debts, the coefficients of having a second job are highly significant in both models; however, they are different in magnitude. Also, having a dependent child is significant once weights based on waves 1, 5, 10 and 15 are used to estimate the model.

4. DISCUSSION

The substantive comparison between the models in this paper shows that using ordinary longitudinal non-response weights to analyse wealth data from waves 5, 10 and 15 from the BHPS does not take into account 478 respondents who have actually provided data usable for analysis in this combination of waves. Compared to a weighting strategy that is designed specifically to consider these 478 respondents, the ordinary weighting strategy provides different results.

Weights from a single weighting strategy do take care of a part of non-response error on several estimates, but clearly fail in tackling the error introduced in other estimates due to the loss of information.

In longitudinal surveys non-response is not a one-off event, it is rather dynamic and can take different patterns among different sub-periods of time during the life of the panel. Therefore, non-response error can vary not just between survey estimates but also within and between sub-periods of times for the same estimate in the same survey. Consequently, different combinations of data collection points might suffer from different sizes of non-response error. This variation might be due to changes in the sample size and/or the sample composition among different combinations of waves. Thus, an ordinary weighting strategy, which does not take into account the changes in the responding sample between wave-combinations, can only deal with the fixed part of non-response error. Instead, a subset of weights that takes into account the change in the responding sample can tackle the fixed as well as the variable part of non-response error.

The consideration of wave-combinations that have the same module of questions as a criterion to design subsets of weights evidently showed an impact on estimates. Hence, this can be considered as a more adequate strategy. However, other features of longitudinal surveys may push for different types of considerations to be taken into account too. For example, to enhance the accuracy of survey estimates, survey organisations sometimes add extra information to the original sample. For instance, two samples (from Scotland & Wales) were added to BHPS in wave 9. Thus, for BHPS, providing subsets of weights for wave 9 onwards might be of interest too.

Table a.1. Logistic regression models of possession of savings and debt.

	Having Savings		**Having Debts**	
	Using weights based on waves 1, 5, 10, and 15	**Using weights based on all waves up to wave 15**	**Using weights based on waves 1, 5, 10, and 15**	**Using weights based on all waves up to wave 15**
Year 2000	0.057	0.080	0.037	0.060
Year 2005	0.025	0.039	0.130**	0.116**
Female	0.208***	0.201***	0.134**	0.147**
Age	-0.007***	-0.008***	-0.040***	-0.041***
Financially okay	-0.841***	-0.832***	0.436***	0.468***
Having financial deficits	-2.384***	-2.439***	1.400***	1.401***
Mortgage payer	-0.117*	-0.122*	1.025***	0.995***
Council tenant	-0.416***	-0.422***	0.961***	0.948***
Private renter	-0.477***	-0.509***	0.922***	0.886***
Having a second job	-0.246***	-0.089	-0.466***	-0.242***
Having a dependent child	-0.302***	-0.297***	0.121*	0.105
Living with partner	0.001	-0.012	0.024	0.062
Member of a large household	-0.473**	-0.478**	-0.282	-0.247
Unemployed	-0.116*	-0.152	-0.234***	-0.270***
Out of the labour force	-1.129***	-1.121***	-0.963***	-0.992***
Annual income/1000	0.018***	0.019***	0.004*	0.004*
Constant	0.894***	0.933***	0.172	0.183
N	5132	4654	5132	4654

Note: * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

Bibliografy

Gouskova E. (2001) *PSID* Technical Report: The 2002 PSID Child Development Supplement (CDS-II) Weights, *[online] availablefrom*http://psidonline.isr.umich.edu/data/weights/ [Accessed 19th December 2010].

Kroh M. (2009) *Data Documentation: Sample Sizes and Panel Attrition in the German Socio Economic Panel (1984 until 2008),* Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).

Lynn P., Kaminska O. (2010) Criteria for Developing Non-response Weight Adjustments for Secondary Users of Complex Longitudinal Surveys, *2010 International Workshop on Household Survey Non-Response.* Nurnberg.

Plaza S., Graf E. (2008) *Executive Summary: Recommendations and Practical Examples for Using Weighting,* Lausanne: University of Lausanne.

Taylor M. (ed.). (2010) *British Household Panel Survey User Manual Volume A*. Colchester: Institute for Social and Economic Research, University of Essex.

Appendix A.2: Effect of the imputation of the amounts of savings and debt.

Table a.2 The distribution of the imputed and un-imputed amounts of savings and debts.

	Mean	Std.dev	Min	Q1	Median	Q3	Max	Skewness	Kurtosis	N%
Saving	67	186	0	0	0	60	5,000	9.9	176	80.1%
Imputed saving	67	182	0	0	0	65	5,000	9.4	165.6	100%
Debt	1,430	6,250	0	0	0	500	400,000	28.2	1365.8	92.8%
Imputed debt	1,489	6,845	0	0	0	600	400,000	31.2	1532.7	100%

* The results indicate that the imputed variables have similar distributions to the un-imputed variables.

Chapter 2 appendices

Appendix B.1: Males death registration in England and Wales from 1992 to 2008.

Table b.1 Male registered deaths by single-year age (1992 to 2008).

Age	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
16	128	118	119	109	134	157	153	129	135	112	138	121	112	115	129	116	122
17	205	205	190	192	208	211	178	198	176	189	192	175	177	193	193	173	174
18	257	244	193	247	230	261	273	234	258	263	251	251	224	226	212	193	198
19	282	260	227	252	225	242	271	264	250	280	235	220	218	246	231	219	213
20	285	318	292	278	241	281	241	239	258	256	277	275	257	231	243	229	266
21	333	318	276	324	267	310	251	240	256	250	252	253	212	248	229	252	222
22	337	333	318	290	284	271	284	246	259	248	278	279	262	242	262	241	245
23	316	307	314	330	295	290	302	266	279	229	249	286	274	260	241	243	264
24	350	321	331	333	318	316	326	274	276	293	254	255	252	236	237	253	258
25	316	345	329	337	348	375	322	297	286	288	276	266	268	281	276	255	236
26	355	342	376	350	313	346	371	334	287	281	282	257	268	248	253	280	287
27	365	381	349	391	339	318	380	364	368	324	298	242	263	248	288	269	277
28	388	360	373	398	392	344	386	384	360	320	330	307	275	212	238	294	303
29	377	416	358	415	400	406	393	400	399	394	349	335	299	280	283	262	287
30	382	374	399	413	429	380	429	409	394	361	406	363	299	329	303	298	289
31	400	426	436	448	469	428	423	396	401	418	396	399	360	311	323	322	307
32	373	357	409	443	458	414	470	458	456	418	435	426	428	365	382	370	300
33	385	427	464	449	433	444	493	463	442	491	434	428	414	433	378	382	378
34	410	411	455	455	446	460	444	473	456	535	459	455	407	439	408	406	424
35	432	437	426	437	482	415	492	513	499	493	474	483	490	450	467	461	496
36	433	443	456	476	457	408	492	490	552	505	536	510	532	515	494	487	468
37	436	436	461	484	470	485	515	495	595	520	577	555	541	562	520	498	498
38	559	510	468	516	488	513	497	544	534	581	541	556	550	534	550	526	600
39	553	534	527	526	586	557	556	559	586	623	604	614	625	581	595	607	581
40	554	580	548	560	613	570	550	592	620	663	644	674	652	643	619	646	659
41	636	621	618	646	602	607	629	653	669	621	671	706	610	724	704	656	742
42	659	666	652	683	656	663	629	651	653	700	687	734	729	728	736	804	727
43	773	709	727	763	741	721	719	676	677	728	776	813	810	822	774	773	823
44	917	836	827	769	766	779	807	745	750	750	745	795	821	803	856	806	825
45	1,016	975	850	887	924	811	880	827	783	828	900	797	826	899	881	885	929
46	949	1,101	1,017	991	981	879	908	958	879	904	912	931	911	927	927	911	975
47	1,072	1,052	1,176	1,124	1,062	1,080	1,049	993	1,005	1,010	1,020	1,019	974	990	1,010	954	1,072
48	1,156	1,183	1,040	1,359	1,244	1,162	1,113	1,055	1,059	1,074	1,051	1,075	1,047	1,109	1,087	1,026	1,065

* Continued in the next page. Source: Office for National Statistics (www.ons.gov.uk).

Table b.1 (continued)

Age	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
49	1,214	1,239	1,235	1,258	1,493	1,359	1,216	1,226	1,253	1,227	1,226	1,133	1,150	1,107	1,135	1,132	1,159
50	1,262	1,360	1,310	1,410	1,289	1,567	1,419	1,396	1,473	1,252	1,306	1,317	1,197	1,247	1,207	1,156	1,297
51	1,300	1,414	1,456	1,448	1,416	1,482	1,682	1,611	1,474	1,440	1,353	1,366	1,387	1,363	1,328	1,261	1,322
52	1,559	1,469	1,522	1,564	1,663	1,584	1,641	1,863	1,713	1,630	1,554	1,481	1,532	1,424	1,403	1,340	1,360
53	1,749	1,746	1,546	1,662	1,696	1,776	1,746	1,790	1,994	1,861	1,730	1,670	1,591	1,462	1,596	1,568	1,455
54	1,840	1,890	1,729	1,696	1,778	1,865	1,952	1,914	1,722	2,198	1,959	1,908	1,802	1,630	1,682	1,660	1,635
55	2,104	2,055	1,972	1,932	1,867	1,845	2,051	2,025	1,982	1,957	2,314	2,300	1,932	1,897	1,834	1,857	1,778
56	2,237	2,246	2,123	2,192	2,154	1,993	2,052	2,151	2,287	2,197	2,095	2,513	2,241	2,048	2,109	2,014	1,875
57	2,453	2,589	2,456	2,386	2,294	2,371	2,108	2,344	2,288	2,340	2,318	2,416	2,686	2,459	2,298	2,098	2,095
58	2,736	2,731	2,549	2,580	2,572	2,539	2,550	2,296	2,245	2,491	2,547	2,467	2,328	2,764	2,646	2,501	2,227
59	2,984	2,889	2,865	2,797	2,807	2,772	2,714	2,621	2,423	2,505	2,650	2,672	2,525	2,578	2,925	2,825	2,608
60	3,422	3,280	2,998	3,144	3,085	3,009	2,946	2,948	2,834	2,580	2,734	2,930	2,829	2,888	2,777	3,065	2,997
61	3,835	3,726	3,468	3,248	3,326	3,283	3,243	3,155	3,188	2,976	2,718	2,892	2,991	2,991	3,033	2,992	3,364
62	4,229	4,136	3,862	3,709	3,559	3,468	3,487	3,415	3,347	3,316	3,185	2,993	3,139	3,169	3,220	3,236	3,108
63	4,615	4,531	4,266	4,141	3,974	3,683	3,740	3,651	3,628	3,519	3,453	3,363	2,951	3,219	3,362	3,371	3,439
64	5,059	5,085	4,765	4,698	4,473	4,147	4,056	3,926	3,781	3,718	3,793	3,745	3,495	3,279	3,347	3,549	3,602
65	5,497	5,478	5,222	5,046	4,993	4,683	4,417	4,286	4,192	3,981	4,057	3,943	3,864	3,647	3,412	3,596	3,714
66	6,191	6,057	5,558	5,605	5,292	5,048	4,906	4,706	4,399	4,307	4,238	4,208	4,180	3,939	3,824	3,597	3,689
67	6,700	6,781	6,231	6,034	5,701	5,642	5,465	5,262	4,871	4,605	4,584	4,570	4,382	4,333	4,188	3,981	3,729
68	7,176	7,072	6,784	6,455	6,270	6,066	5,829	5,666	5,332	4,895	4,868	4,906	4,677	4,555	4,515	4,471	4,346
69	7,780	7,656	7,021	7,160	6,798	6,446	6,219	6,180	5,893	5,525	5,309	5,090	5,024	4,775	4,647	4,793	4,688
70	8,563	8,113	7,696	7,588	7,569	7,027	6,829	6,581	6,206	6,035	5,721	5,361	5,246	5,039	5,001	4,825	4,981
71	9,583	9,253	8,195	8,061	7,825	7,673	7,287	7,132	6,611	6,483	6,420	6,057	5,529	5,556	5,331	5,153	5,290
72	9,948	10,018	8,892	8,612	8,458	8,026	7,934	7,603	7,111	6,840	6,784	6,478	6,241	5,840	5,604	5,541	5,562
73	7,464	10,472	10,226	9,548	8,820	8,495	8,225	8,226	7,784	7,171	7,031	6,999	6,635	6,353	5,887	5,699	5,963
74	7,455	7,836	10,315	10,415	9,733	9,169	8,878	8,375	8,402	7,796	7,572	7,452	6,931	6,982	6,472	6,174	5,900
75	8,178	7,676	7,618	10,693	10,673	9,727	9,064	8,930	8,510	8,440	8,127	7,717	7,436	7,025	7,010	6,782	6,506
76	9,282	8,667	7,437	8,004	10,777	10,712	9,915	9,241	8,766	8,651	8,647	8,229	7,636	7,647	7,352	7,354	7,015
77	10,095	9,491	8,387	7,633	7,722	10,860	10,542	10,095	9,216	8,967	8,864	8,658	8,152	7,918	7,624	7,584	7,517
78	10,244	10,288	9,268	8,775	7,611	7,779	10,867	10,771	10,007	9,216	8,986	8,723	8,662	8,203	7,845	7,774	7,772
79	9,897	10,571	9,754	9,527	8,507	7,515	7,881	10,919	10,468	9,877	9,461	9,371	8,775	8,593	8,360	8,001	8,210
80	9,875	10,259	9,959	9,871	9,404	8,399	7,424	7,571	10,467	10,435	9,954	9,368	8,896	8,702	8,602	8,278	8,236
81	9,323	9,844	9,522	9,877	9,615	8,795	8,100	7,221	7,475	10,163	10,620	10,009	8,908	8,994	8,539	8,682	8,635
82	9,155	9,345	9,131	9,737	9,460	9,292	8,605	8,095	7,025	7,309	10,269	10,443	9,475	8,809	8,821	8,790	8,631
83	8,650	8,986	8,596	9,124	9,035	9,077	9,015	8,471	7,603	6,898	7,103	10,215	9,785	9,227	8,779	8,677	8,586
84	7,963	8,590	8,323	8,376	8,360	8,503	8,635	8,649	7,912	7,282	6,462	6,863	9,642	9,543	8,980	8,651	8,691
85	7,094	7,916	7,457	7,736	7,900	7,982	8,049	8,139	7,891	7,644	6,962	6,624	6,356	8,902	9,092	8,710	8,491
86	6,317	6,912	6,628	7,200	7,181	7,133	7,281	7,393	7,478	7,368	7,067	6,594	5,890	6,191	8,653	8,744	8,671
87	5,576	5,989	5,853	6,315	6,464	6,624	6,531	6,650	6,736	6,660	6,864	6,640	6,029	5,484	5,577	8,119	8,262
88	4,676	5,214	5,046	5,429	5,609	5,759	5,759	6,025	5,940	6,203	6,354	6,300	5,846	5,533	4,816	5,262	7,699
89	3,933	4,353	4,275	4,547	4,807	5,010	5,083	5,228	5,078	5,379	5,644	5,782	5,433	5,461	4,961	4,599	4,943
90+	9,165	14,087	13,818	15,418	16,117	16,800	17,510	18,522	19,020	19,579	20,818	22,124	21,775	22,976	23,083	23,496	23,810

* Source: Office for National Statistics (www.ons.gov.uk).

Appendix B.2: Female death registration in England and Wales from 1992 to 2008.

Table b.2 Female registered deaths by single-year age (1992 to 2008).

Age	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
16	68	80	59	72	80	75	99	66	65	81	79	64	79	72	62	65	47
17	95	83	74	76	84	92	92	97	69	83	84	76	72	80	84	83	71
18	104	82	74	82	90	91	97	90	118	86	86	83	87	90	91	85	90
19	108	98	91	92	99	87	92	86	107	77	99	98	113	94	88	81	94
20	120	120	84	84	97	84	88	93	101	92	84	92	89	96	88	75	86
21	127	116	106	101	108	103	80	105	88	104	97	90	98	85	83	99	94
22	121	105	114	108	90	96	101	110	98	96	87	108	95	94	88	89	92
23	122	103	110	107	111	92	96	73	105	92	97	103	114	107	86	83	92
24	125	134	113	121	111	128	101	103	112	88	102	92	98	96	99	107	94
25	131	125	117	117	136	125	137	95	116	100	103	98	110	106	107	88	98
26	137	159	135	133	146	119	123	129	116	128	105	123	115	114	128	113	120
27	124	184	153	190	157	119	142	129	136	126	108	112	103	128	110	94	111
28	176	174	141	152	169	131	152	178	140	124	118	134	116	112	129	116	120
29	172	154	179	161	165	186	163	139	156	154	142	155	132	127	111	132	137
30	157	189	183	184	192	170	195	186	174	152	153	176	167	127	133	130	134
31	196	197	195	177	182	213	198	181	195	197	196	174	169	152	129	126	140
32	206	209	227	231	236	203	190	190	195	216	201	193	186	181	140	165	171
33	205	189	230	250	239	214	245	238	227	210	216	210	204	207	193	179	198
34	218	228	212	247	260	231	223	242	247	258	255	261	234	227	202	217	216
35	256	216	248	271	262	283	265	269	286	315	251	249	257	214	234	236	215
36	253	269	270	234	302	303	280	301	316	309	276	285	277	269	265	229	224
37	268	270	289	268	283	315	300	301	300	301	300	288	309	308	289	276	289
38	289	258	287	317	336	322	325	302	334	350	340	359	364	316	308	311	325
39	328	313	337	340	328	355	340	352	343	354	364	359	372	359	370	355	358
40	370	357	378	358	354	356	375	366	394	369	376	420	374	395	387	422	393
41	379	436	407	404	397	360	426	420	418	414	378	419	430	396	417	441	440
42	486	443	438	424	433	425	395	459	459	424	446	447	452	485	462	476	472
43	470	508	487	464	494	485	449	484	492	535	497	501	488	502	514	523	532
44	618	541	538	562	483	530	550	519	511	487	539	557	532	561	556	518	563
45	693	691	637	614	566	546	561	561	541	548	571	551	563	537	594	614	597
46	666	756	752	680	634	597	586	591	630	614	643	623	640	571	573	599	675
47	685	717	853	781	720	760	640	663	680	642	664	704	688	660	665	666	659
48	735	733	768	909	817	818	740	758	695	709	754	723	696	757	762	746	727

* Continued in the next page. Source: Office for National Statistics (www.ons.gov.uk).

Table b.2 (continued)

Age	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
49	752	763	809	872	921	882	881	822	806	787	785	770	717	794	748	729	744
50	808	907	913	925	949	1,054	1,035	1,002	919	916	823	872	855	832	855	840	907
51	795	887	875	1,040	1,038	991	1,181	1,073	975	984	896	919	908	878	840	861	939
52	980	954	922	1,043	1,045	1,085	1,066	1,222	1,234	1,093	1,018	1,014	877	982	967	869	915
53	1,079	1,044	1,043	1,052	1,090	1,130	1,119	1,077	1,355	1,198	1,170	1,128	1,047	1,020	1,022	1,026	961
54	1,143	1,150	1,106	1,071	1,061	1,192	1,268	1,230	1,273	1,493	1,365	1,250	1,148	1,144	1,073	1,122	1,167
55	1,218	1,248	1,193	1,274	1,168	1,153	1,294	1,302	1,339	1,302	1,530	1,518	1,264	1,252	1,186	1,215	1,178
56	1,322	1,332	1,288	1,368	1,262	1,200	1,245	1,340	1,429	1,410	1,440	1,657	1,531	1,429	1,409	1,290	1,230
57	1,530	1,518	1,438	1,418	1,492	1,438	1,346	1,393	1,481	1,515	1,534	1,411	1,760	1,708	1,456	1,448	1,345
58	1,636	1,633	1,463	1,607	1,617	1,575	1,535	1,419	1,460	1,546	1,560	1,617	1,570	1,856	1,699	1,590	1,499
59	1,794	1,821	1,710	1,686	1,733	1,715	1,704	1,678	1,547	1,579	1,639	1,782	1,687	1,672	2,039	1,832	1,803
60	2,074	2,006	1,833	1,830	1,837	1,928	1,864	1,902	1,774	1,632	1,730	1,871	1,858	1,719	1,735	2,124	2,055
61	2,369	2,324	2,081	1,986	2,009	2,069	1,968	1,998	1,919	1,929	1,770	1,853	1,848	1,956	1,958	2,028	2,380
62	2,624	2,619	2,336	2,226	2,171	2,213	2,158	2,134	2,069	2,029	2,013	1,830	1,913	2,035	2,100	1,996	2,047
63	2,830	2,916	2,690	2,551	2,304	2,246	2,259	2,297	2,184	2,304	2,181	2,091	1,985	2,036	2,170	2,324	2,211
64	3,188	3,046	2,942	2,864	2,721	2,516	2,499	2,486	2,520	2,362	2,410	2,371	2,233	2,134	2,196	2,319	2,439
65	3,603	3,488	3,261	3,369	3,132	2,990	2,687	2,680	2,608	2,596	2,480	2,642	2,420	2,401	2,191	2,248	2,499
66	3,843	3,864	3,627	3,528	3,336	3,404	3,106	2,917	2,862	2,720	2,828	2,682	2,717	2,772	2,521	2,380	2,392
67	4,231	4,241	3,987	3,713	3,778	3,675	3,604	3,397	3,109	2,948	2,959	2,975	2,850	2,932	2,803	2,699	2,568
68	4,814	4,592	4,424	4,332	4,197	4,034	3,848	3,718	3,470	3,299	3,198	3,191	3,213	3,013	3,127	2,885	2,913
69	5,190	5,111	4,813	4,765	4,664	4,285	4,349	4,141	3,999	3,667	3,469	3,473	3,371	3,331	3,105	3,312	3,198
70	5,882	5,792	5,367	5,273	5,213	4,879	4,711	4,600	4,343	4,224	3,867	3,743	3,526	3,486	3,485	3,558	3,492
71	6,692	6,482	5,871	5,769	5,552	5,485	5,261	4,993	4,762	4,675	4,562	4,202	3,865	3,736	3,748	3,709	3,692
72	7,451	7,349	6,694	6,370	6,193	6,096	5,915	5,554	5,088	5,048	5,045	4,886	4,460	4,301	4,004	3,991	4,004
73	5,757	8,082	7,566	7,299	6,819	6,651	6,331	6,164	5,604	5,522	5,432	5,380	5,038	4,628	4,436	4,344	4,441
74	5,960	6,261	8,168	8,248	7,578	7,054	6,930	6,794	6,473	5,940	5,805	5,827	5,581	5,313	5,082	4,777	4,684
75	6,996	6,271	6,324	8,795	8,624	8,108	7,335	7,254	6,897	6,701	6,617	6,343	6,015	5,887	5,642	5,262	5,054
76	7,990	7,515	6,370	6,855	9,305	9,125	8,535	8,054	7,543	7,229	7,190	6,937	6,456	6,389	6,028	6,020	5,782
77	9,011	8,692	7,570	6,797	7,249	9,807	9,528	8,884	8,275	7,849	7,681	7,650	7,022	6,767	6,559	6,544	6,327
78	9,569	9,724	8,666	8,105	7,391	7,543	10,217	10,212	9,073	8,414	8,451	8,171	7,796	7,470	7,060	6,858	6,956
79	10,066	10,465	9,651	9,359	8,610	7,536	7,761	10,590	10,222	9,569	9,010	8,992	8,059	8,194	7,791	7,539	7,445
80	10,494	10,981	10,588	10,537	9,784	8,922	7,817	8,220	10,872	10,765	10,244	9,721	8,870	8,633	8,253	8,193	8,040
81	10,879	11,250	10,808	11,152	10,993	9,969	9,180	8,151	8,188	11,474	11,397	10,719	9,691	9,445	8,935	8,808	8,771
82	11,211	11,665	11,155	11,373	11,582	11,458	10,431	9,646	8,287	8,609	12,038	12,172	10,686	10,058	9,431	9,398	9,222
83	11,570	12,123	11,319	11,514	11,929	11,783	11,479	10,644	9,647	8,518	8,839	12,576	11,695	11,261	10,230	10,065	9,817
84	11,762	12,234	11,544	11,929	11,849	11,758	11,998	11,705	10,647	9,908	8,746	9,413	12,114	12,205	11,281	10,724	10,413
85	11,441	12,223	11,703	11,919	11,843	11,827	11,683	12,295	11,677	10,705	10,101	9,277	9,117	12,407	12,224	11,637	11,024
86	11,245	11,787	11,466	11,935	12,059	11,817	11,771	12,040	11,723	11,435	10,827	10,429	8,732	9,016	12,180	12,527	11,694
87	10,895	11,593	10,970	11,484	11,707	11,543	11,460	11,649	11,465	11,642	11,580	11,218	9,659	8,663	8,837	12,297	12,668
88	9,923	10,891	10,457	11,017	11,208	11,364	11,259	11,201	11,064	11,141	11,590	11,586	10,276	9,604	8,336	8,697	12,730
89	9,215	10,212	9,621	10,137	10,414	10,876	10,795	10,912	10,422	10,741	11,074	11,440	10,511	10,120	8,907	8,161	8,836
90+	42,751	48,212	46,465	50,987	52,117	54,062	55,682	58,390	57,462	58,488	61,277	64,018	61,193	63,060	61,384	62,433	62,875

* Source is Office for National Statistics (www.ons.gov.uk).

Appendix B.3: Male resident population in England and Wales from 1992 to 2008.

Table b.3 Male residents by single-year age (1992 to 2008). *Thousands*

Age	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
16	310.8	322.7	318.9	326.8	326.7	324.9	329	330.6	327.9	343.7	343.3	350.1	360.8	358.5	361.3	367.9	360.4
17	315	326.9	323.1	331	330.9	329.1	333.2	330.6	330.7	330.6	347.1	346.9	353.9	365.6	361.3	364.7	370.6
18	318.2	330.1	326.3	334.2	334.1	332.3	336.4	334.9	330.6	321.2	334	352.2	352.4	359.3	369.6	366	367.9
19	307.4	319.3	315.5	323.4	323.3	321.5	325.6	337.6	335.2	313.9	325.2	341.9	359.5	359.8	366.8	376.8	370.4
20	284	295.9	292.1	300	299.9	298.1	302.2	326.1	340.4	325.2	315.8	329.9	346.3	365.5	363.5	374.2	380.8
21	280.6	292.5	288.7	296.6	296.5	294.7	298.8	304.6	328.7	325.7	327.4	320.5	335.3	353.8	373.6	373	379.7
22	294	305.9	302.1	310	309.9	308.1	312.2	302.4	308.2	312.8	328.4	333.6	326.9	344.4	367.5	384.6	379.6
23	308.5	320.4	316.6	324.5	324.4	322.6	326.7	317.0	306.9	297.4	316.6	336.3	340.6	336.8	358.3	379.3	391.9
24	321.3	333.2	329.4	337.3	337.2	335.4	339.5	332.4	322.2	297.5	301.6	325.6	343.5	350.8	349.1	370.1	386.5
25	347.4	359.3	355.5	363.4	363.3	361.5	365.6	346.1	340.6	318.8	301.2	311.8	331	349.7	360.1	356.6	376.8
26	371.6	383.5	379.7	387.6	387.5	385.7	389.8	371.5	353.3	334.4	321.8	311.6	316.8	336.9	359.3	367	363
27	391.2	403.1	399.3	407.2	407.1	405.3	409.4	395.5	378.3	345.8	336.9	324.2	316.2	322.1	341.7	365.6	373
28	390.1	402	398.2	406.1	406	404.2	408.3	414.7	402.1	368.5	348	337.9	328.8	321.4	327.6	348	371.4
29	402.6	414.5	410.7	418.6	418.5	416.7	420.8	413.1	420.9	388.1	370.2	348.7	341.8	333.4	329.5	333.1	353.2
30	403.5	415.4	411.6	419.5	419.4	417.6	421.7	424.9	418.5	398.6	389.1	368.5	348.7	343.0	337.9	329	336.6
31	415.3	427.2	423.4	431.3	431.2	429.4	433.5	425.2	429.7	396.1	399.5	387.2	368.5	349.6	347.1	337.2	331.7
32	423.7	435.6	431.8	439.7	439.6	437.8	441.9	436.3	428.7	406.0	396.7	398.0	387.1	369.4	354.5	346.4	339.8
33	431.7	443.6	439.8	447.7	447.6	445.8	449.9	444.2	439.8	409.6	406.9	395.2	397.7	387.6	372.6	353.8	348.8
34	430.9	442.8	439	446.9	446.8	445	449.1	451.8	453.4	420.4	410.4	405.0	394.9	398.2	386.7	371.8	355.8
35	420.1	432	428.2	436.1	436	434.2	438.3	450.7	452.2	412.4	419.9	408.6	405.5	396.9	397.4	388.3	355.8
36	407.8	419.7	415.9	423.8	423.7	421.9	426	439.8	441.1	413.0	411.4	418.5	409	407.2	396.4	398.7	355.8
37	396.5	408.4	404.6	412.5	412.4	410.6	414.7	427.3	428.5	410.1	411.7	417.9	418.9	410.6	405.9	397.6	398.2
38	375.6	387.5	383.7	391.6	391.5	389.7	393.8	416.2	417.3	404.0	409.1	418.8	418.1	420.4	411	406.8	397
39	362.3	374.2	370.4	378.3	378.2	376.4	380.5	395.3	396	394.9	403.1	414.4	418.9	419.2	422	411.8	397
40	353.2	365.1	361.3	369.2	369.1	367.3	371.4	381.3	382.1	385.6	394.5	406.6	413.7	418.5	419.3	421.6	411.2
41	339.8	351.7	347.9	355.8	355.7	353.9	358	372.2	372.7	372.1	385.7	396.4	405.7	413.0	419.4	421.6	420.9
42	327.1	339	335.2	343.1	343	341.2	345.3	359.6	360.4	365.0	372.6	386.6	395.6	405.1	412.7	418.7	417.9
43	316.7	328.6	324.8	332.7	332.6	330.8	334.9	345.9	346.1	360.3	365.8	372.7	385.6	394.7	402.7	412	417.8
44	321.1	333	329.2	337.1	337	335.2	339.3	335.2	334.9	348.7	361	365.5	371.7	384.8	392.8	402.1	411
45	318.8	330.7	326.9	334.8	334.7	332.9	337	339.3	338.9	336.3	347.2	323.9	364	371.0	383.7	390.2	401.3
46	311.2	323.1	319.3	327.2	327.1	325.3	329.4	336.9	336.4	327.7	335	320.8	357.7	363.2	383.6	381.3	389.4
47	318	329.9	326.1	334	333.9	332.1	336.2	329.2	328.7	329.9	326.1	322.0	344.2	356.8	362.5	366.3	380.4
48	326	337.9	334.1	342	341.9	340.1	344.2	335.8	335.1	326.8	328.2	331.4	331.6	343.3	357.8	360.1	365.4

* Continued in the next page. Source: statistics.gov.uk.

Table b.3 (continued) *Thousands*

Age	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
49	338	349.9	346.1	354	353.9	352.1	356.2	343.7	343.1	323.3	325.4	342.6	323	330.6	342.4	355.5	359.2
50	361.5	373.4	369.6	377.5	377.4	375.6	379.7	355.2	354.5	324.3	321.8	323.9	325.1	321.7	329.1	340.7	354.7
51	397	408.9	405.1	413	412.9	411.1	415.2	378.6	377.7	333.9	322.9	406.6	322.2	323.8	320.7	327.3	339.9
52	302	313.9	310.1	318	317.9	316.1	320.2	413.9	412.4	345.4	332.5	396.4	318.9	320.7	321.5	318.8	326.4
53	296.7	308.6	304.8	312.7	312.6	310.8	314.9	319.0	317.7	371.6	344.1	386.6	320.1	317.4	317.6	319.6	317.9
54	298.1	310	306.2	314.1	314	312.2	316.3	313.5	311.7	392.4	368.8	372.7	329.2	318.5	315.1	319.6	318.4
55	280.6	292.5	288.7	296.6	296.5	294.7	298.8	314.7	312.7	319.1	390.7	365.5	339.7	326.8	316.2	312.7	313
56	253.7	265.6	261.8	269.7	269.6	267.8	271.9	297.0	294.8	312.1	317.5	359.2	364.7	336.9	324.1	313.4	310
57	230.8	242.7	238.9	246.8	246.7	244.9	249	270.0	267.8	305.2	309.4	345.8	383.9	361.7	324.1	321.4	310.6
58	244.4	256.3	252.5	260.4	260.3	258.5	262.6	247.0	267.8	291.2	302.8	333.2	311.7	380.6	359	331.9	318.6
59	244.8	256.7	252.9	260.8	260.7	258.9	263	260.2	257.6	261.5	288.5	324.6	304.2	308.7	376.6	355.7	328.7
60	241.3	253.2	249.4	257.3	257.2	255.4	259.5	260.3	257.4	244.7	259	326.7	296.1	300.0	304.4	370.4	352.7
61	234.4	246.3	242.5	250.4	250.3	248.5	252.6	256.6	253.4	254.3	242.1	255.5	281.3	291.9	295.9	298.9	366.9
62	227.1	239	235.2	243.1	243	241.2	245.3	249.4	246.1	254.2	251.3	238.9	251.5	277.1	288.1	290.6	295.7
63	220.4	232.3	228.5	236.4	236.3	234.5	238.6	241.9	238.2	251.6	251.1	247.6	234.9	247.3	274	282.6	287.3
64	209.7	221.6	217.8	225.7	225.6	223.8	227.9	234.8	230.7	244.6	248.2	247.1	243.2	230.8	244.8	268.6	279.1
65	205.6	217.5	213.7	221.6	221.5	219.7	223.8	224.0	219.5	235.8	241	244.3	242.8	239.3	227.8	240.6	264.7
66	208.6	220.5	216.7	224.6	224.5	222.7	226.8	219.3	214.7	227.8	231.9	236.8	239.7	238.7	235.2	223.9	236.8
67	207	218.9	215.1	223	222.9	221.1	225.2	221.7	216.4	216.2	223.6	227.8	232.1	235.2	233.6	230.9	220.1
68	199.1	211	207.2	215.1	215	213.2	217.3	219.7	213.9	212.3	211.9	219.0	222.8	227.4	228.8	229.1	226.6
69	188.7	200.6	196.8	204.7	204.6	202.8	206.9	211.4	205.1	210.7	207.4	206.9	213.7	218.0	220.6	223.8	224.5
70	181.5	193.4	189.6	197.5	197.4	195.6	199.7	200.6	194	207.1	205.1	202.5	201.5	208.3	211.4	215.6	218.8
71	177.5	189.4	185.6	193.5	193.4	191.6	195.7	192.8	185.7	198.3	200.9	199.8	196.8	195.7	201.8	206.1	210.4
72	173.8	185.7	181.9	189.8	189.7	187.9	192	188.2	180.4	188.3	191.9	194.9	193.5	190.9	188.8	196.2	200.7
73	162	173.9	170.1	178	177.9	176.1	180.2	184.1	175.6	178.3	181.4	185.4	188.1	186.9	183.7	183.1	190.5
74	153.7	165.6	161.8	169.7	169.6	167.8	171.9	171.8	163.1	173.0	171.1	174.6	178.2	181.3	179.2	177.6	177.2
75	145.3	157.2	153.4	161.3	161.2	159.4	163.5	162.9	154	165.3	165	163.8	167.1	171.2	173.2	172.6	171.3
76	143.4	155.3	151.5	159.4	159.3	157.5	161.6	154.4	145	154.7	157	157.3	156	159.5	162.9	166.1	165.6
77	139.1	151	147.2	155.1	155	153.2	157.3	151.6	141.3	145.0	145.8	148.5	149	148.1	151.3	155.3	158.7
78	134.2	146.1	142.3	150.2	150.1	148.3	152.4	146.6	135.9	136.0	136	137.3	139.9	140.6	139.8	143.5	147.7
79	75.4	87.3	83.5	91.4	91.3	89.5	93.6	141.0	129.7	131.7	126.9	127.1	128.4	131.1	132.2	132	135.7
80	63.4	75.3	71.5	79.4	79.3	77.5	81.6	86.0	129.7	126.8	122	117.9	117.8	119.7	122.4	123.9	123.8
81	65	76.9	73.1	81	80.9	79.1	83.2	74.4	67.3	115.9	116.4	112.4	108.5	108.7	111	113.9	123.8
82	62.3	74.2	70.4	78.3	78.2	76.4	80.5	75.1	67.1	78.5	106	106.7	102.6	99.5	99.8	102.3	105.2
83	55.9	67.8	64	71.9	71.8	70	74.1	71.9	63.6	60.4	71.6	95.0	96.7	93.1	90.5	91.1	93.7
84	47.2	59.1	55.3	63.2	63.1	61.3	65.4	65.3	63.6	59.1	54	65.0	84.6	87.0	83.8	81.9	82.6
85	37.8	49.7	45.9	53.8	53.7	51.9	56	56.8	56.8	54.0	52	47.8	58.6	74.9	77.5	75.1	73.3
86	28.1	40	36.2	44.1	44	42.2	46.3	48.2	48.9	49.2	46.8	45.6	41.7	52.3	65.6	68.6	66.5
87	20.5	32.4	28.6	36.5	36.4	34.6	38.7	39.4	41	41.4	42.2	40.1	39.3	36.1	46.2	56.9	60.4
88	13	24.9	21.1	29	28.9	27.1	31.2	32.3	33	34.0	35	35.9	33.8	33.4	30.7	40.9	48.6
89	7.1	19	15.2	23.1	23	21.2	25.3	25.5	26.8	27.0	28.1	29.1	30	28.2	28	26.1	36
90	53.7	65.6	61.8	69.7	69.6	67.8	71.9	76.3	80	77.0	81	85.1	89.5	93.9	95.9	98.5	98.7

* Source: statistics.gov.uk.

Appendix B.4: Female resident population in England and Wales from 1992 to 2008.

Table b.4 Female residents by single-year age (1992 to 2008). *Thousands*

Age	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
16	293.2	304.1	300.3	308.2	302.1	306.3	310.4	311.7	310	323.7	324.3	331.7	341	338.7	339.7	344.5	341
17	299.3	310.2	306.4	314.3	308.2	312.4	316.5	311	311	312.6	325.9	327.6	334.7	344.9	340	342.2	347.6
18	302	312.9	309.1	317	310.9	315.1	319.2	317.7	309.9	307.4	315.2	330.3	331.8	340.4	347	344.6	347.7
19	290.7	301.6	297.8	305.7	299.6	303.8	307.9	320.6	315.4	308.7	311.6	321.7	335.9	339.9	343.5	354.5	353.5
20	269.4	280.3	276.5	284.4	278.3	282.5	286.6	307.9	323.3	322.9	313.1	316.9	327.2	343.4	350	349.8	360.6
21	266.2	277.1	273.3	281.2	275.1	279.3	283.4	288.4	310	325	328	318.5	323.4	336.3	355	357.7	357.5
22	280.6	291.5	287.7	295.6	289.5	293.7	297.8	286.9	292.1	315	331.5	335.2	326.4	334.1	347.9	364.1	366.9
23	294.9	305.8	302	309.9	303.8	308	312.1	302.5	291.9	303.1	322	338.3	342.7	336.3	343.9	356.4	372.6
24	307.7	318.6	314.8	322.7	316.6	320.8	324.9	317.1	307.1	305.8	310.3	327.9	345.6	352.5	348.9	352.2	364.5
25	329.8	340.7	336.9	344.8	338.7	342.9	347	330.1	324.1	315.5	309.9	312.3	332.2	352.7	357.8	353.2	353.2
26	351.1	362	358.2	366.1	360	364.2	368.3	351.2	335.9	329.4	318.9	312.1	316.3	339.0	357.8	361.9	354.1
27	372.3	383.2	379.4	387.3	381.2	385.4	389.5	372.1	356.3	343.2	332.8	320.9	315.5	322.1	344.3	361.4	362.6
28	367.5	378.4	374.6	382.5	376.4	380.6	384.7	392.6	377.1	364.8	346.4	334.5	324	320.9	328.9	347.5	362.1
29	382.8	393.7	389.9	397.8	391.7	395.9	400	387.3	396.4	386.9	367.9	347.8	337.4	328.7	327.1	331.7	348.1
30	385.6	396.5	392.7	400.6	394.5	398.7	402.8	402.5	390.9	398.4	389.8	369.5	349	340.2	335.4	326.3	333.8
31	397.1	408	404.2	412.1	406	410.2	414.3	404.6	405.1	394.6	401.2	392.0	370.6	351.3	346.4	334.6	328.1
32	405.3	416.2	412.4	420.3	414.2	418.4	422.5	416.1	407.1	406.4	397	403.4	392.9	372.8	355.4	345.6	336.2
33	412.4	423.3	419.5	427.4	421.3	425.5	429.6	423.8	418.2	409.8	408.6	399.1	404	394.7	374.6	354.5	346.9
34	411.4	422.3	418.5	426.4	420.3	424.5	428.6	430.7	425.7	418.8	411.6	410.5	399.9	405.4	395.7	373.8	355.7
35	402.9	413.8	410	417.9	411.8	416	420.1	429.7	432.1	421	419.6	412.3	410.9	400.7	405.4	396.2	374.2
36	392.8	403.7	399.9	407.8	401.7	405.9	410	421.3	431.3	423	421.2	420.7	412.7	411.7	399.7	405.8	396.4
37	380.9	391.8	388	395.9	389.8	394	398.1	410.9	422.4	419.4	422.9	422.7	421.1	413.3	411.8	400.1	406
38	362.9	373.8	370	377.9	371.8	376	380.1	399	412	412.4	419.3	424.6	423	421.6	413.3	412.1	400.3
39	356.5	367.4	363.6	371.5	365.4	369.6	373.7	380.9	400.2	402.2	412.3	420.8	424.8	423.3	422	413.5	412.1
40	348.3	359.2	355.4	363.3	357.2	361.4	365.5	374.2	382	391.3	402.1	413.9	420.6	424.8	422.8	422.8	412.6
41	337.2	348.1	344.3	352.2	346.1	350.3	354.4	366	375.2	375.8	391.3	403.4	413.7	420.6	424.3	423.5	421.8
42	324.4	335.3	331.5	339.4	333.3	337.5	341.6	354.9	366.9	368.4	375.6	392.0	403.1	413.6	420	425	422.5
43	314.2	325.1	321.3	329.2	323.1	327.3	331.4	341.9	355.5	363.2	368.5	376.4	391.7	403.0	412.6	420.4	424
44	319.9	330.8	327	334.9	328.8	333	337.1	331.8	342.7	352.1	363.2	369.0	376	391.5	401.8	412.9	419.5
45	317.9	328.8	325	332.9	326.8	331	335.1	337.3	332.3	340.2	351.4	363.4	368.1	375.5	390.4	401.4	412.8
46	312.6	323.5	319.7	327.6	321.5	325.7	329.8	335.1	337.5	332.4	339.9	351.3	362.6	367.6	374.3	390.1	401.2
47	317.2	328.1	324.3	332.2	326.1	330.3	334.4	329.6	335.4	334.9	331.5	339.6	350.4	362.0	366.1	373.9	389.8
48	327.9	338.8	335	342.9	336.8	341	345.1	334.2	329.8	331.3	333.9	331.5	338.6	349.8	360.9	365.5	373.6

* Continued in the next page. Source: statistics.gov.uk.

Table b.4 (continued) *Thousands*

Age	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
49	339.1	350	346.2	354.1	348	352.2	356.3	344.8	334.1	328.5	330.4	334.0	330.5	337.9	348.5	360.3	365.1
50	363.9	374.8	371	378.9	372.8	377	381.1	355.7	344.7	330.5	327.7	329.5	332.7	329.3	336.4	347.1	359.1
51	396.8	407.7	403.9	411.8	405.7	409.9	414	380.3	355.5	341	329.5	326.9	328.1	331.4	327.9	335	346
52	305.3	316.2	312.4	320.3	314.2	318.4	322.5	413	379.8	351.3	340	328.7	325.5	326.7	329.8	326.5	333.8
53	300.7	311.6	307.8	315.7	309.6	313.8	317.9	321.7	412.3	377.9	350.4	339.2	327.2	324.2	324.8	328.3	325.2
54	301.1	312	308.2	316.1	310	314.2	318.3	317	321.1	397.4	375.4	349.2	337.5	325.6	322.6	323.1	327
55	284	294.9	291.1	299	292.9	297.1	301.2	317.3	316.2	324.8	396.7	375.6	347.3	335.7	324.1	321.5	320.9
56	257.2	268.1	264.3	272.2	266.1	270.3	274.4	300.1	316.2	317.7	324	394.7	373.4	345.3	334.2	322.8	319.2
57	235.4	246.3	242.5	250.4	244.3	248.5	252.6	273.3	298.9	309.8	316	322.4	392.3	371.1	343.6	332.7	320.5
58	249.5	260.4	256.6	264.5	258.4	262.6	266.7	251.5	272	296.8	308.5	315.1	320.5	389.9	368.6	342	330.4
59	252.1	263	259.2	267.1	261	265.2	269.3	265.2	250.1	268.6	295.2	306.9	312.9	318.4	386.9	366.8	339.4
60	248.7	259.6	255.8	263.7	257.6	261.8	265.9	267.7	263.9	252.8	267.2	293.3	304.9	311.4	315.2	384	365
61	241.8	252.7	248.9	256.8	250.7	254.9	259	264.1	266.2	263	251	264.8	291.3	303.2	307.3	312.6	381.7
62	237.2	248.1	244.3	252.2	246.1	250.3	254.4	257	262.5	263.9	260.9	248.8	262.8	289.5	298.8	304.6	310.9
63	230.1	241	237.2	245.1	239	243.2	247.3	252.3	255.2	260.9	261.8	258.7	246.8	260.9	286.6	296	302.6
64	222.1	233	229.2	237.1	231	235.2	239.3	245.1	250.2	255.1	258.7	259.3	256.4	244.9	258.5	283.8	293.8
65	221.6	232.5	228.7	236.6	230.5	234.7	238.8	236.9	243.1	248.9	252.7	256.1	256.7	253.5	241.5	255.6	281.7
66	225.5	236.4	232.6	240.5	234.4	238.6	242.7	236.2	234.5	241.7	246.4	250.1	253.4	253.7	251.5	238.9	253.4
67	230.9	241.8	238	245.9	239.8	244	248.1	239.6	233.5	233.5	239.1	243.5	247.3	250.2	250.8	248.2	236.8
68	229.1	240	236.2	244.1	238	242.2	246.3	244.6	236.5	233.5	230.9	236.0	240.3	243.9	247.1	247.4	245.7
69	221.6	232.5	228.7	236.6	230.5	234.7	238.8	242.4	241.1	235.1	230.4	227.3	232.6	236.7	240.5	243.5	244.5
70	215.6	226.5	222.7	230.6	224.5	228.7	232.8	234.5	238.2	237.3	231.4	226.6	223.7	229.0	233.2	236.8	240.3
71	214.7	225.6	221.8	229.7	223.6	227.8	231.9	228.1	229.9	233.2	233	227.4	222.8	219.9	225.1	229.3	233.4
72	214	224.9	221.1	229	222.9	227.1	231.2	226.5	223	225.7	228.6	228.7	222.8	218.5	216.1	221	225.7
73	209.3	220.2	216.4	224.3	218.2	222.4	226.5	225.1	221.1	217.4	220.6	223.7	223.6	218.0	214	211.7	216.9
74	206.5	217.4	213.6	221.5	215.4	219.6	223.7	220	218.8	216.1	211.9	215.4	218.2	218.3	213.1	209.2	207.4
75	204	214.9	211.1	219	212.9	217.1	221.2	216.6	213	212.8	210.1	206.0	209.3	212.4	212.8	207.7	204.4
76	206.9	217.8	214	221.9	215.8	220	224.1	213.4	209.1	206.7	206	203.6	199.7	202.9	206.3	206.7	202.3
77	212	222.9	219.1	227	220.9	225.1	229.2	215.3	205.1	201.9	199.6	199.1	196.5	193.0	196.2	199.9	200.7
78	212.7	223.6	219.8	227.7	221.6	225.8	229.9	219.5	206.3	197.3	194.1	192.0	191.4	189.3	185.7	189.4	193.2
79	131.3	142.2	138.4	146.3	140.2	144.4	148.5	218.9	209.2	199.4	189.1	185.7	183.7	183.2	181.5	178.4	182.4
80	119.2	130.1	126.3	134.2	128.1	132.3	136.4	140.9	207.5	201.2	189.9	180.0	176.9	175.2	175	173.5	170.7
81	127	137.9	134.1	142	135.9	140.1	144.2	128.5	132.7	188.9	189.7	180.0	170.4	167.6	166.2	166.5	165.2
82	128.9	139.8	136	143.9	137.8	142	146.1	134.8	120.1	134.8	178	179.1	169.2	160.5	157.8	157	157.4
83	127.1	138	134.2	142.1	136	140.2	144.3	135.5	125.2	110.6	126.2	164.7	167.1	158.2	150	148.3	147.5
84	116.5	127.4	123.6	131.5	125.4	129.6	133.7	132.5	124.8	114.1	101.9	117.9	151.9	155.2	146.4	139.5	138.1
85	101.5	112.4	108.6	116.5	110.4	114.6	118.7	121.3	120.7	110.7	104.2	93.3	108.9	138.7	142.7	135.1	128.8
86	87.6	98.5	94.7	102.6	96.5	100.7	104.8	106.7	109.2	108.1	99.9	94.2	84.2	99.8	125.7	130.5	123.5
87	76.5	87.4	83.6	91.5	85.4	89.6	93.7	92.9	94.8	96.9	96.4	89.2	84.1	75.6	90.9	112.8	118.2
88	64.8	75.7	71.9	79.8	73.7	77.9	82	82	81.6	84.2	85.3	85.0	78.4	74.3	67.1	82.5	99.8
89	55.1	66	62.2	70.1	64	68.2	72.3	70.8	71.2	70.9	73	74.0	73.8	67.9	65.2	58.9	73.9
90	258.5	269.4	265.6	273.5	267.4	271.6	275.7	283.8	290.9	262.7	269.1	275.4	280.1	284.8	286.2	286.4	279

* Source: statistics.gov.uk.

Appendix B.5: 95% CIs of the difference between equivalent coefficients resulting from modelling subjective health status.

Table b.5 Full results of the random effects logistic regression models of the determinants of subjective health status.

	Respondents aged 16 to 59			Respondents aged 60+			All respondents aged 16+		
	Using SWs	Using AWs_1	Using AWs_2	Using SWs	Using AWs_1	Using AWs_2	Using SWs	Using AWs_1	Using AWs_2
Year 1997 to 2002	0.941(*.038)*	0.938(*.035)* (-0.098-0.104)	0.938(*.035)* (-0.098-0.104)	0.993(*.091)*	1.037(*.070)* (-0.269-0.181)	1.078(*.061)* (-0.300-0.130)	0.950(*.035)*	0.961(*.032)* (-0.104-0.082)	0.982(*.030)* (-0.122-0.058)
Year 2003 to 2008	0.910(*.037)***	0.895(*.034)*** (-0.083-0.113)	0.893(*.033)*** (-0.080-0.114)	1.018(*.094)*	1.048(*.072)* (-0.262-202)	1.033(*.058)* (-0.231-0.201)	0.928(*.034)**	0.931(*.031)** (-0.093-0.087)	0.942(*.029)** (-0.102-0.074)
Female	1.110(*.073)*	1.128(*.070)* (-0.216-0.180)	1.128(*.070)* (-0.216-0.180)	0.558(*.100)***	0.621(*.083)**** (-0.318-0.192)	0.507(*.055)**** (-0.173-0.275)	0.991(*.062)*	0.981(*.055)** (-0.152-0.172)	0.893(*.047)** (-0.054-0.250)
White	1.685(*.207)****	1.767(*.201)**** (-0.648-0.484)	1.790(*.203)**** (-0.673-0.463)	1.183(*.673)*	1.692(*.345)* (-1.991-0.973)	1.855(*.327)* (-2.139-0.795)	1.677(*.205)****	1.729(*.194)**** (-0.605-0.501)	1.728(*.191)**** (-0.600-0.498)
Age	0.991(*.002)****	0.990(*.002)**** (-.005-0.007)	0.990(*.002)**** (-0.005-0.007)	0.988(*.007)*	0.983(*.005)*** (-0.022-0.012)	0.980(*.004)*** (-0.008-0.024)	0.991(*.002)****	0.990(*.001)**** (-0.003-0.005)	0.989(*.001)**** (-0.002-0.006)
1 to 2 visits to GP since last year	0.352(*.033)****	0.357(*.031)**** (-0.094-0.084)	0.357(*.031)**** (-0.094-0.084)	0.639(*.112)**	0.609(*.078)**** (-0.238-0.298)	0.511(*.056)**** (-0.117-0.373)	0.392(*.032)****	0.408(*.029)**** (-0.101-0.069)	0.396(*.027)**** (-0.086-0.078)
3 to 5 visits to GP since last year	0.087(*.008)****	0.087(*.008)**** (-0.022-0.022)	0.087(*.007)**** (-0.021-0.021)	0.252(*.045)****	0.246(*.032)**** (-0.102-0.114)	0.202(*.022)**** (-0.048-0.148)	0.104(*.009)****	0.112(*.008)**** (-0.032-0.016)	0.113(*.008)**** (-0.033-0.015)
6 + visits to GP since last year	0.016(*.001)****	0.016(*.001)**** (-0.003-0.003)	0.016(*.001)**** (-0.003-0.003)	0.087(*.015)****	0.084(*.011)**** (-0.033-0.039)	0.073(*.008)**** (-0.019-0.047)	0.021(*.002)****	0.024(*.002)**** (-0.009-0.003)	0.025(*.002)**** (-0.010-0.002)
Smoker	0.696(*.039)****	0.699(*.037)**** (-0.108-0.102)	0.700(*.037)**** (-0.109-0.101)	0.639(*.110)***	0.647(*.083)**** (-0.278-0.262)	0.635(*.072)**** (-0.254-0.262)	0.686(*.037)****	0.688(*.033)**** (-0.099-0.095)	0.683(*.032)**** (-0.093-0.099)
Annual income/1000	1.007(*.002)****	1.007(*.002)**** (-0.006-0.006)	1.007(*.002)**** (-0.006-0.006)	0.979(*.006)***	0.976(*.004)**** (-0.011-0.017)	0.975(*.004)**** (-0.010-0.018)	1.002(*.003)*	1.003(*.002)** (-0.008-0.006)	1.005(*.002)** (-0.010-0.004)
Has a partner	0.983(*.123)*	0.989(*.099)* (-0.315-0.303)	0.925(*.077)* (-0.226-0.342)	1.016(*.054)**	1.019(*.051)** (-0.149-0.143)	1.020(*.050)** (-0.148-0.140)	1.027(*.050)*	1.042(*.045)* (-0.147-0.117)	1.039(*.042)* (-0.140-0.116)
Financially okay	0.860(*.045)***	0.872(*.043)*** (-0.134-0.110)	0.874(*.042)*** (-0.135-0.107)	0.727(*.082)***	0.707(*.061)**** (-0.180-0.220)	0.626(*.045)**** (-0.082-0.284)	0.840(*.040)****	0.838(*.035)**** (-0.102-0.106)	0.795(*.032)**** (-0.055-0.145)
Financially struggling	0.568(*.031)****	0.576(*.029)**** (-0.091-0.075)	0.577(*.029)**** (-0.092-0.074)	0.613(*.072)****	0.598(*.053)**** (-0.160-0190)	0.457(*.033)**** (0.00-0.311)	0.569(*.028)****	0.575(*.025)**** (-0.080-0.068)	0.532(*.022)**** (-0.033-0.107)
N	3,594	3,594	3,594	503	503	503	4,097	4,097	4,097
σ	1.60	1.60	1.61	1.69	1.72	1.72	1.62	1.62	1.62
ρ	0.44	0.44	0.44	0.46	0.47	0.47	0.44	0.45	0.45

* Entries are odds ratios, their standard errors and 95% CIs of the difference between the coefficients adjusted with AWs_1 or AWs_2 and the equivalent coefficients adjusted with *SWs*. All CIs include a zero indicating no significant difference between the equivalent coefficients estimated with the *SWs and AWs.* The reference categories of the categorical independent variables are: year 1991 to 1996, male, non-white, no visits to the GP since last year, non-smoker, does not have a partner and having good financial situation. . σ is the standard error of the random effects (sigma u). ρ is the percentage of the total variance that is due to differences between units. * p<0.05, ** p<0.01, *** p<0.001.