Reporting bias and heterogeneity in self-assessed health. Evidence from the British Household Panel Survey

Cristina Hernandez-Quevedo[1], Andrew M Jones[1], Nigel Rice[2]

March 2004

Abstract

This paper explores reporting bias and heterogeneity in the measure of self-assessed health (SAH) used in the British Household Panel Survey (BHPS). The ninth wave of the BHPS includes the SF-36 general health questionnaire, which incorporates a different wording to the self-assessed health variable used at other waves. Considerable attention has been devoted to the reliability of SAH and the scope for contamination by measurement error; the change in wording at wave 9 provides a form of natural experiment that allows us to assess the sensitivity of panel data analyses to a change in the measurement instrument. In particular, we investigate reporting bias due explicitly to the change in the question. We show how progressively more general specifications of reporting bias can be implemented using panel data ordered probit and generalised ordered probit models. Our results suggest that the distribution of SAH does shift at the ninth wave but there is little evidence that this varies with socio-economic characteristics at an individual level.

JEL codes I12 C23

Keywords: self-assessed health, reporting bias, ordered probit, generalised ordered probit, panel data

---

Data from the *British Household Panel Survey* (BHPS) were supplied by the ESRC Data Archive. Neither the original collectors of the data nor the Archive bear any responsibility for the analysis or interpretations presented here. This paper derives from the project "The dynamics of income, health and inequality over the lifecycle" (known as the ECuity III Project), which is funded in part by the European Community's Quality of Life and Management of Living Resources programme (contract QLK6-CT-2002-02297). Nigel Rice is in part supported by the UK Department of Health programme of research at the Centre for Health Economics. The authors would like to thank Xander Koolman his helpful advice and participants in the York Seminars in Health Econometrics (YSHE) for their comments.

[1]Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, United Kingdom: Tel: +44-1904-433790; Fax: +44-1904-433759; Email: amj1@york.ac.uk
[2]Centre for Health Economics, University of York, United Kingdom

1. Introduction

Like many general population surveys, the British Household Panel Survey (BHPS) includes a measure of self-assessed health (SAH): respondents are asked to rate their own health on a five-point categorical scale. The current release of the BHPS spans eleven years, 1991-2001, and provides the opportunity for longitudinal analysis of the relationship between self-assessed health and socio-economic status, allowing in-depth investigation of dynamics (state dependence) and individual heterogeneity. Longitudinal analyses of SAH face a challenge – the SF-36 questionnaire was included in wave 9 of the BHPS and, with it, the SAH question was re-worded and included a modification to the response categories. Does this allow satisfactory multivariate analyses of SAH that exploit the full span of the longitudinal data? With this challenge there is also an opportunity. Considerable attention has been devoted to the reliability of SAH and the scope for contamination by measurement error; the change in wording at wave 9 provides a form of natural experiment that allows us to assess the sensitivity of panel data analyses to a change in the measurement instrument. We investigate whether different groups of individuals respond to the change in measurement instrument in different ways.

The structure of the paper is as follows. The self-assessed health variable in the BHPS is introduced in Section 2. Section 3 describes the issue of reporting bias and the estimation and testing strategy. The BHPS dataset is described in Section 4. Results are shown in Section 5, concluding the paper in Section 6.

2. Self-assessed health in the BHPS

Self-assessed health is often included in general socio-economic surveys, such as the BHPS and the European Community Household Panel (ECHP). SAH has been used widely in previous studies of the relationship between health and socioeconomic status (e.g., Adams et al. (2003), Benzeval et al. (2000), Deaton and Paxson (1998), Ettner (1996), Frijters et al. (2003), Salas (2002), Smith (1999)) and of the relationship between health and lifestyles (e.g., Kenkel (1995), Contoyannis and Jones (2003)). SAH is a simple subjective measure of health that provides an ordinal ranking of perceived health status. However it has been shown to be a powerful predictor of subsequent mortality (see e.g., Idler and Kasl (1995), Idler and Benyamini (1997)) and its predictive power does not appear to vary across socioeconomic groups (see e.g., Burström and Fredlund (2001)). Categorical measures of SAH have been shown to be good predictors of subsequent use of medical care (see e.g., van Doorslaer et al. (2000), van Doorslaer, Koolman and Jones (2002)). Socioeconomic inequalities in SAH have been a focus of research (see e.g., van Doorslaer et al. (1997), van Doorslaer and Koolman (2002), van Ourti (2003)) and have been shown to predict inequalities in mortality (see e.g., van Doorslaer and Gerdtham (2003)).

This subjective measure of health has caused debate in the literature concerning its validity. It has been argued that perceived health does not correspond with actual health (see Bound, 1990), while other researchers have regarded this variable as a valid indicator of true health (see Butler *et al*., 1987). As a self-reported subjective measure of health, SAH may be prone to measurement error. General evidence of non-random measurement error in self-reported health is reviewed in Currie and Madrian (1999). Crossley and Kennedy (2002) report evidence of measurement error in a 5-category SAH question. They exploit the fact that a random sub-sample of respondents to the 1995 Australian National Health Survey were asked the same version of the SAH question twice, before and after other morbidity questions. The first question was administered as part of the SF-36 questionnaire on a self-completion form, the second as part of a face-to-face interview on the main questionnaire. They find a statistically significant difference in the distribution of SAH between the two questions and evidence that these differences are related to age, income and occupation. This measurement error could be explained by a *mode of administration effect*, due to the use of self-completion and face-to-face interviews (Grootendorst et al. (1997) find evidence that self-completion questions reveal more morbidity); or a *framing or learning effect* by which SAH responses are influenced by the intervening morbidity questions.

It is sometimes argued that the mapping of "true health" into SAH categories may vary with respondent characteristics. This source of measurement error has been termed "state-dependent reporting bias" (Kerkhofs and Lindeboom (1995)), "scale of reference bias" (Groot (2000)) and "response category cut-point shift" (Sadana *et al.* (2000), Murray *et al*. (2001)). This occurs if sub-groups of the population use systematically different cut-point levels when reporting their SAH, despite having the same level of true health.

Regression analysis of SAH can be achieved through specifying an ordered probability model, such as the ordered probit or logit. In the context of ordered probit models the symptoms of measurement error can be captured by making the cut points dependent on some or all of the exogenous variables used in the model and estimating a generalised ordered probit. This requires strong *a priori* restrictions on which variables affect health and which affect reporting in order to separately identify the influence of variables on latent health and on measurement error. It is worth noting that allowing the scaling of SAH to vary across individual characteristics is equivalent to a heteroskedastic specification of the underlying latent variable equation (see e.g., van Doorslaer and Jones (2003)). This is because location and scale cannot be separately identified in binary and ordered choice models and, in general, it is not possible to separate measurement error from heterogeneity. Attempts to surmount this problem include modelling the reporting bias based on more "objective" indicators of true health (Kerkhofs and Lindeboom (1995), Lindeboom and van Doorslaer (2003)) and the use of "vignettes" to fix the scale (Murray *et al*. (2001)). Lindeboom and van Doorslaer (2003) analyse SAH in the Canadian National Population Health Survey and use the McMaster Health Utility Index (HUI-3) as their objective measure of health. They find evidence of cut-point shift with respect to age and gender, but not for income, education or linguistic group.

In general, the SAH question is measured following an ordinal scale, with possible responses from "very poor" or "poor" to "very good" or "excellent". The SAH variable that is included in the BHPS has two wordings, depending on the wave that is taken into account. For waves 1-8 and 10-11, the SAH variable represents "health status over the last 12 months". Respondents are asked: "Compared to people of your own age, would you say your health over the last 12 months on the whole has been: excellent, good, fair, poor, very poor?". However, the SF-36 questionnaire was included in wave 9. The SF-36 questionnaire includes 36 items that measure health across eight dimensions of physical functioning, social functioning, role limitations due to physical problems, role limitations due to emotional problems, mental health, vitality, pain, and general health perceptions (Jenkinson *et al*., 1996). In this questionnaire, the SAH variable for wave 9 represents "general state of health", using the question: "In general, would you say your health is: excellent, very good, good, fair, poor?".

If the two wordings of the question are compared, it is possible to distinguish two main differences. Firstly, the self-assessed health question in wave 9 does not include the age benchmark that is present in the rest of the waves, in which individuals are asked to assess their level of health "compared to people of their own age". The second difference is the way in which the categories have been labelled. Although both questions provide 5 possible answers to the respondents, the category "very poor" is not available in wave 9, but "very good" is incorporated between "good" and "excellent".

Figure 1 presents the distribution of SAH for each wave, using a balanced panel of individuals who are observed for all 11 waves. Accordingly, all wave-specific distributions are based on the same sample of individuals followed over time. The different categories are shown on the horizontal axis of each graph, with "1" representing the lowest level of health[1], and "5", the highest level[2]. The histograms for waves 1-8 and 10-11 follow a similar pattern; a skewed distribution is clear, with the majority of respondents reporting their health as "good" or "excellent". Around 50 per cent of individuals report a "good" level of health and around 70 per cent, either "good" or "excellent" levels of health. These percentages vary little across waves. Note that although the question is framed in terms of self-rating of health compared to people of ones own age, the distribution of responses across the ten waves indicates a worsening of reported health as the cohort ages. For example, 5% of individuals report "poor" health in wave 1 compared to 7.5% in wave 11; 30% of respondents report "excellent" health in wave 1 compared to 21% in wave 11.

At wave 9, where the SF-36 version of SAH was used there is a notable change in the distribution of responses. Fewer individuals report their health status within the two highest categories ("excellent" and "very good" health), while a greater proportion of respondents report within the bottom two categories ("poor" and "fair" health). In general, there appears to be a shift to the left of the distribution together with a flattening of its mass point. There is also evidence of a floor effect: individuals who report their health as "poor" in wave 9, may have assessed their health as "very poor" if this category had been available.

3. Multivariate models for ordered responses in longitudinal data

3.1 Reporting bias

Given the change in the distribution of the SAH variable at wave 9, reporting bias could be considered as an explanation for this observation. Reporting bias has been a concern in the literature. The systematic use of different threshold levels by sub-groups of a population reflect the existence of reporting bias (Lindeboom and van Doorslaer, 2003; Murray *et al*, 2001; Groot, 2000; Sadana *et al*, 2000; Kerkhofs and Lindeboom, 1995). These differences may be influenced by, among other things, age, gender, education, income, language and personal experience of illness. Basically, it means that different groups appear to interpret the question within their own specific context and therefore use different reference points when they are responding to the same question.

Lindeboom and van Doorslaer (2003) distinguish between *index shift* and *cut-point shift*. *Index shift* occurs if the shape of the distribution of SAH remains the same, but there is a change in its location such that there is a parallel shift in all of the reporting thresholds for particular sub-groups of the population. This is illustrated in Figure 2:

[1] This corresponds to "very poor" health at waves 1-8, 10-11 and "poor" health at wave 9.
[2] Corresponding to "excellent" health at all waves.

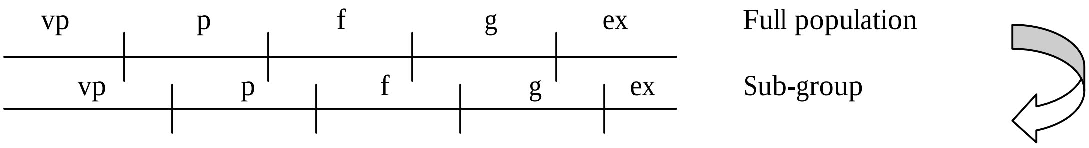

Figure 2

It should be stressed that the term "index shift" may be misleading. The parallel shift in the distribution may be due to a shift in the cut-points (reporting bias) or due to a shift in the underlying measure of "true health". In general, it is not possible to separately identify the two reasons for index shift.

If the reporting bias is due to *cut-point shift*, this implies that there is a change in the relative positions of the reporting thresholds for particular sub-groups of the population (see Figure 3 for an example). This may or may not result in a change in the overall distribution of SAH.

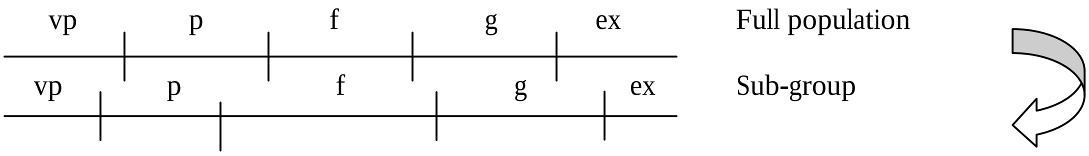

Figure 3

A comparison of the two versions of the SAH question in the BHPS has the added complexity of the wave 9 SF-36 version of SAH having a different wording and a different set of response categories compared to the wave 1-8 and 10-11 question. The resulting differences in the distributions if SAH at wave 9 compared to other waves could be due to either an index shift in the mapping of "true" health to the SAH categories or a change in the relative location of the cut-points. Figure 4 illustrates a cut-point shift at wave 9.

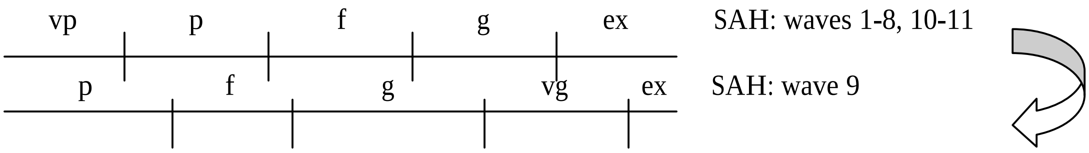

Figure 4

Figure 1 shows evidence that, at least at an aggregate level, individuals appear to retain a reasonably consistent interpretation of the original SAH question. That is, the distribution of responses to the SAH question over waves 1-8 and 10-11 are very similar. However, there is a trend towards reporting worsening health over the eleven waves. This may be indicative of "true" health deteriorating as the cohort ages or it could be due to a shift in reporting thresholds used by respondents. The distribution of SAH at wave 9 clearly differs from the distributions observed at other waves. The joint effect of a change in the framing of the question together with the addition of the "very good" category and the removal of the "very poor" category is the likely source of the majority of the differences observed.[3] It is of interest to ascertain whether these effects have resulted in either an index shift or a cut-point shift in the relationship mapping latent "true" health to the observed categories of SAH.

3.2 Pooled ordered probit and random effects ordered probit

One of the aims of this paper is to test whether the change in the measurement of SAH at wave 9 of the BHPS influences the estimates of panel data models for ordered responses. For that purpose, pooled ordered probit (POP) and random effects ordered probit (REOP) models have been applied. The literature suggests that ordered probit or ordered logit models can be used when considering an ordered categorical variable, such as SAH (McElvey and Zavoina, 1975), and if the scaling of the variable is available, interval regression is an alternative (van Doorslaer and Jones, 2003). In this paper we follow Contoyannis et al (2003) and apply POP and REOP. These models are easily implemented using STATA v8 (Frechette, 2001). The pooled model is estimated with robust standard errors, allowing

[3] It is possible that the distribution of "true" health at wave 9 differs from the distribution of "true" health at other waves and that this is reflected in self-assessed health status irrespective of the framing of the question. However in light of the evidence on the consistency in the reported distribution of SAH across the other waves this would appear highly unlikely.

for clustering of observations within individuals (see e.g., Wooldridge, 2002). In the case of the REOP, the log-likelihood for each unit is approximated by Gauss-Hermite quadrature (Butler & Moffit, 1982).

Ordered probit (and logit) models are typically motivated by a latent variable specification: this makes an explicit distinction between the process that determines an individuals "true" health – the *latent variable model* – and the thresholds that determine which category of SAH they report, given their true health – the *measurement model*. The *latent variable* specification for the models considered in this study is given by:

$$h_{it}^* = \eta' x_{it} + \text{Ǭ}_{it} \qquad i=1,\dots,N;\ t=1,\dots,T \qquad (1)$$

where Ǭ_{it} is an idiosyncratic error term uncorrelated with the set of regressors ζ_{it}. h_{it}^* represents the "true" level of health for individual i at time t. As "true" health is not observable we rely on the use of SAH as an appropriate dependent variable. In the particular case of the REOP, Ǭ_{it} is regarded as a two-component error term such that, $\text{Ǭ}_{it} = \bar{\bar{a}}_i + \text{ǣ}_{it}$,. Here, $\bar{\bar{a}}_i$ represents an individual-specific time-invariant component and ǣ_{it} is an idiosyncratic component, uncorrelated over time (Verbeek, 2000). ζ_i and κ_{it} are assumed to be mutually independent and independent of x_{it}. For maximum likelihood estimation we assume normality of the errors.

It is useful to note that consistency of the POP estimator relies on the assumption that x_{it} is independent of the latent error, Ǭ_{it}, in each time period. Stronger exogeneity assumptions are required for consistency of the REOP estimator. Here it is assumed that x_{it} is independent of $\text{Ǭ}_{it} = \bar{\bar{a}}_i + \text{ǣ}_{it}$ for all time periods. Accordingly, x_{it} should be independent of $\bar{\bar{a}}_i$ and past, current and future values of ǣ_{it} (see Wooldridge, 2002, p405 & p490). This strict exogeneity assumption of the REOP estimator places additional restrictions on the specification of the model that are not required for the POP estimation.

Direct comparison of the coefficients from the POP and REOP is not possible because of a different scaling of the error variance: the POP gives *population-averaged* coefficients, while the REOP gives *cluster-specific* coefficients(Zeger et al., 1988). The POP assumes that the error term, Ǭ_{it}, is distributed N(0,1) for identification of η, while the REOP assumes that ǣ_{it} only is N(0,1) and, hence, the overall error variance equals $(\omega^2_\zeta + 1)$. The cluster-specific ηs estimated for the REOP can be re-scaled to give population-averaged coefficients using $\eta / \cdot (1+\omega^2_\zeta)$ (see e.g., Wooldridge, 2002, p.486).

The *measurement model* relates values of h_{it}^* to reported categories of SAH. If the re-wording of the SAH variable at wave 9 is taken into account, the latent variable specification would be the following[4]:

- For the original wording of the question, i.e., for waves 1-8 and 10-11,

  $h_{it} = j$ *if* $\sigma^0_{j-1} < h^*_{it} < \sigma^0_j$, with j=1 (very poor), 2 (poor), 3(fair), 4(good), 5(excellent)
- In the case of the re-wording of the question (wave 9),

  $h_{it} = j$ *if* $\sigma^1_{j-1} < h^*_{it} < \sigma^1_j$, with j = 1 (poor), 2 (fair), 3(good), 4 (very good), 5 (excellent)

Given this change in the phrasing of the question, the probability of an individual i choosing the j[th] category will be given by the following[5]:

$$P(h_{it} = j \mid x_{it}, \zeta_i) =$$
$$A(\sigma^0_j - \eta' x_{it} - \zeta_i) - A(\sigma^0_{j-1} - \eta' x_{it} - \zeta_i) \qquad t = 1\text{-}8,\ 10\text{-}11 \qquad (2)$$
$$A(\sigma^1_j - \eta' x_{it} - \zeta_i) - A(\sigma^1_{j-1} - \eta' x_{it} - \zeta_i) \qquad t = 9$$

It can be seen that the η's and ζ_i coefficients are common for all waves; given the latent variable specification, these parameters should be unaffected by changes in the measurement model. This assumption provides the basis of our tests for reporting bias. Interest focuses on the estimates of the cut-points σ^0_j and σ^1_j .

3.3. Empirical Tests

The simplest form of index shift is where there is a parallel shift in the reporting thresholds that is common to all individuals. This implies that the original cut-points (σ^0_j) can be obtained by adding a constant, $\bar{a}$, to the cut-points corresponding to wave 9 (σ^1_j):

---

[4] To identify the model, we follow STATA in using the conventional normalisation, that is: σ^0_0=-Ð, σ^0_5=+Ð and there is no constant term in the linear index given by equation (1)

[5] Here we are assuming a REOP specification with an individual-specific error component, ζ_i

$\sigma^{0}_{j} = \sigma^{1}_{j} + \nu,$ *for all j* (3)

In this particular case, the probability that an individual i chooses category j to assess their health in time t, can be written as:

$P(h_{it} = j \mid x_{it}, \zeta_i) =$

$$A(\sigma^{0}_{j} - \eta' x_{it} - \zeta_i) - A(\sigma^{0}_{j-1} - \eta' x_{it} - \zeta_i) \qquad t = 1\text{-}8,\ 10\text{-}11$$
(4)
$$A(\sigma^{0}_{j} - \nu - \eta' x_{it} - \zeta_i) - A(\sigma^{0}_{j-1} - \nu - \eta' x_{it} - \zeta_i) \qquad t = 9$$

This shows that the parallel shift in wave 9 can be captured by adding a dummy variable for wave 9 to the latent variable equation – the parallel shift in the cut-points in (3) is equivalent to a shift in the intercept of the latent variable equation (1) at wave 9, hence the label "index shift" used by Lindeboom and van Doorslaer (2003). However, in the context of our application to the BHPS we are maintaining the assumption that the true health equation, (1), does not change at wave 9 but the measurement model, (2), does. So a more accurate label would be "parallel cut-point shift". In practice, time dummies are included for each wave of the panel and the focus is on whether the variable for wave 9 stands out from the underlying trend.

This argument can be extended to the case where the index shift differs across different types of individual, defined by their characteristics *x* measured at wave 9:

$\sigma^{0}_{j} = \sigma^{1}_{j} + \iota' x_{i9},$ *for all j* (5)

Following the same argument as above, the model can be reformulated so that the latent variable is expressed as follows:

$h_{it}^* = \eta' x_{it} + \zeta_i + \kappa_{it},$ $t = 1\text{-}8,\ 10\text{-}11$ (6)
$h_{it}^* = (\eta + \iota)' x_{it} + \zeta_i + \kappa_{it},$ $t = 9$

In this case, differences in reporting at wave 9 are captured by the inclusion of interaction terms between the wave 9 dummy variable and the explanatory variables.

Our specification relies on the distinction between the latent variable model (1) and the measurement model. In particular we have assumed that there is a single index that describes "true" health. This could be relaxed by allowing for a generalised ordered probit specification in which the ηs are allowed to vary across the reported categories of SAH (η_j , j=1…5). This generalised specification makes intuitive sense if the ordered categories reflect choices available to the individual, but this is not the case with self-assessed health. The categories of SAH are an artefact of the design of the survey question and it is hard to construct an argument that the relationship between someone's true health and their socio-economic characteristics will be contingent on the number and labelling of these arbitrary categories. Nevertheless, the existence of individual heterogeneity in the ηs (i.e., allowing for η_i to differ across individuals) such that, for example, the association between income and health is lower for those in better health, could lead to heterogeneity across the categories of SAH (see van Doorslaer and Jones, 2003). To assess the *single index* assumption – sometimes referred to as the proportionality assumption (see Long, 1997, p.140), we estimate the POP and REOP specifications with different numbers of categories of SAH. As well as 5 categories, we also estimate models with SAH collapsed to 4 and 3 categories. Interest focuses on whether the ηs remain stable with different categorisations of the dependent variable.

This approach also sheds light on *cut-point shift*. This occurs when the shift in specific thresholds is allowed to differ. In general it is not possible to distinguish between heterogeneity and measurement error (see e.g., van Doorslaer and Jones, 2003). However, a comparison of specifications with 5, 4 or 3 categories will provide evidence whether the coefficients remain stable across categories.

The previous analysis can be extended, in order to test for *cut-point shift*. For that purpose, we estimate a model allowing parameters to vary across categories of *y* for every wave, following a Generalised Ordered Probit model (GOP). The GOP allows the relaxation of one of the restrictive characteristics of the ordered probit model: the constancy of threshold parameters (see an application of this model in Pudney & Shields, 2000). In order to allow cut-points to vary with the regressors, σ^{0}_{j} needs to be reformulated as follows:

$\mu^{o}_{ij} = \mu^{0}_{j} + \AE^{o}_{j}{}' x_{it}$ *for all j* (7)

If the impact of socio-economic characteristics on the thresholds at wave 9 differs across thresholds, (5) becomes:

$$\sigma^{0}_{ij} = \sigma^{1}_{ij} + \iota^{1}_{j}{}' x_{i9}, \qquad \text{for } t = 9 \tag{8}$$

Then, ι^1_j in equation (8) represents the additional effect of the regressors in the cut-points. Given (1), (7) and (8):

$$P(h_{it} = j \mid x_{it}, \zeta_i) =$$

$$\begin{array}{ll} \Lambda\left(\sigma^{0}_{j} - (\eta - \iota^{0}_{j})' x_{it} - \zeta_i\right) - \Lambda\left(\sigma^{0}_{j-1} - (\eta - \iota^{0}_{j})' x_{it} - \zeta_i\right) & t = 1\text{-}8,\ 10\text{-}11 \\ \Lambda\left(\sigma^{0}_{j} - (\eta + \iota^{1}_{j} - \iota^{0}_{j})' x_{it} - \zeta_i\right) - \Lambda\left(\sigma^{0}_{j} - (\eta + \iota^{1}_{j} - \iota^{0}_{j})' x_{it} - \zeta_i\right) & t = 9 \end{array} \tag{9}$$

This gives a specification in which the coefficients vary across categories of SAH, as a result of cut-point shift. The pooled GOP is estimated directly by maximum likelihood, with robust standard errors corrected for the clustering within-individuals. For identification, the normalisation $\mu^0_{i4} = \mu^0_4$ is imposed, so the cut-point shift associated with the x's is measured relative to μ^0_4.

4. The BHPS data

The BHPS is a longitudinal survey of private households in Great Britain (England, Wales and Scotland south of the Caledonian Canal), and was designed as an annual survey of each adult (16+) member of a nationally representative sample of more that 5,000 households, with a total of approximately 10,000 individual interviews. The first wave of the survey was conducted between 1st September 1990 and 30th April 1991. The initial selection of households for inclusion in the survey was performed using a two-stage stratified systematic sampling procedure designed to give each address an approximately equal probability of selection[6]. The same individuals are re-interviewed in successive waves and, if they split off from their original households are also re-interviewed along with all adult members of their new households.

A balanced panel is used, which means that only individuals from the first wave who were interviewed in each subsequent wave are included. Following Contoyannis et al. (2003), we include individuals who gave a full interview at each wave.[7] Our working sample consists of 2,255 men and 2,841 women. All models are estimated separately for men and women.

Self-assessed health

SAH is defined for waves 1-8 and 10-11 as the response to the question "Compared to people of your own age, would you say your health over the last 12 months on the whole has been: excellent, good, fair, poor, very poor?" At wave 9, where the SF-36 version of the SAH is used, respondents are asked: "In general, would you say your health is: excellent, very good, good, fair, poor?" For both questions the responses are coded in increasing order of health. For example, for waves 1-8 and 10-11, "very poor" health is coded as 1, whilst "excellent" health is coded as 5. At wave 9, "poor" health is coded as 1 and "excellent" health as 5.

Insert Table 1 about here

To assess the single index assumption we collapse the raw categories of SAH so that there is common support within each collapsed category across the two versions of the question. For example, we first collapse SAH into four categories creating a new SAH variable with the following codes: 1: individuals who reported either "very poor" or "poor" health at waves 1-8 and 10-11, or individuals who reported "poor" health at wave 9; 2: individuals who reported "fair" health; 3: individuals who reported "good" health at waves 1-8 and 10-11, or "good" or "very good" health at wave 9; 4: individuals who reported "excellent" health. The original SAH categories were also further collapsed to a three-category variable. The collapsed categorisations are summarised in Table 1.

A further construction of SAH containing 4 categories is performed. This follows an approach suggested by Lindley and Lorgelly (2003). They estimate the relationship between income inequality and self-assessed health using

[6] For further details see Taylor (1998).

[7] This excludes individuals for whom proxy interviews were provided together with those who responded to telephone interviews.

the first nine waves of the BHPS. The interest of their paper for the purpose of this study is the way in which wave 9 data is included in the analysis. In particular, SAH has been constructed as a categorical variable with 4 categories, recoded as "excellent", "good", "fair" and "poor". "Poor" and "very poor" were merged into the new category of "poor". At wave 9 respondents reporting "very good" health were re-assigned to either "good" or "excellent" health on the following basis. A random sample consisting of 33% of individuals was recoded as "excellent". The remaining individuals were recoded as "good". A random sample of 20% of individuals reporting "good" health at wave 9 was recoded as "fair". Further, a random sample of 20% of individuals reporting "fair" health at wave 9 was recoded as "poor". It is argued by the authors that these re-codings and weights were chosen in order to mantain averages for each category of self-assessed health, for all waves (Lindley and Lorgelly, 2003). Table 2, presents the proportion of respondents in each category for the various constructions of the SAH variable. A comparison of the observed cell frequencies across the eleven waves for the four classifications of the SAH variable shows that the approach adopted by Lindley and Lorgelly (2003) provides greatest consistency. In particular, the frequencies for each of the categories observed at wave 9 are close to those obtained for the same categories at other waves.

Insert Table 2 about here

Socio-economic variables

Our empirical models are intended to capture the association between SAH and a range of socio-economic variables. There are three variables representing marital status (Widow, Single, Div/Sep), with Married as the reference category. An indicator of ethnicity is introduced (Non-White). Four dummy variables have been constructed to represent maximum level of education achieved, focusing on the human capital of the respondents. These are: Degree (higher degree/first degree), HND/A (Higher National Certificate, Higher National Diploma, A-level), O/CSE (O-level, Certificate of Secondary Education). No-Qual (no qualifications) is the base case for the education variables. The size of the household (HHSize) and the number of children in the household, by age, are also included in the analysis. The income variable is the logarithm of equivalised real income, adjusted using the Retail Price Index and equivalised by the McClement's scale to adjust for household size and composition. A quartic polynomial function of age is included (Age, Age2=Age^2/100, Age3=Age^3/10,000, Age4 = Age^4/1,000,000). A vector of time dummies is also included in the analysis.

Insert Table 3 about here

5. Results

The latent variable model

Results for POP and REOP models for SAH with 5 categories, estimated for men and women separately on all data excluding wave 9, are presented in Table 4. These results represent a "benchmark" against which the results of incorporating the wave 9 SF-36 version of SAH can be compared. A number of features are worth noting. There is a clear gradient in the relationship between educational attainment and self-assessed health with individuals with greater qualifications reporting better health. There is also the expected positive relationship between income and health, although for both men and women the relationship is more prominent in the POP model. The observed difference between POP and REOP estimators of the income effect is likely to be due to the additional restrictions that the strict exogeneity assumption necessary for consistent estimation using REOP imposes. It is quite plausible that income in period t is correlated with latent health shocks in a previous period, for example, ε_{it-1}. This renders REOP estimates of the relationship between income and health biased and inconsistent. As long as income in period t is not correlated with the idiosyncratic error in the same period, POP remains a consistent estimator of the income effect on health.

Insert Table 4 about here

We observed a general decline in the reporting of better health states when we compared the raw distributions of SAH across the eleven waves as shown in Figure 1. When we condition on relevant regressors such as age, income and educational attainment we still observe a clear gradient across waves, this is evident from the coefficients on the year dummy variables. These become more negative as the cohort ages indicating the reporting of worsening health. This may be due to an age effect not fully removed by the wording of the SAH question or not fully controlled for by conditioning on a polynomial in age. However, it may also be due to changes in reporting as individuals age, reflected in an index shift over time. This would imply that the context within which an individual rates their own health changes with increasing age. However this second hypothesis seems unlikely as the direction of the trend contradicts the notion of *adaptation*: that individuals adapt to worsening health, becoming more optimistic in their assessment of self-assessed health (Groot, 2000).

Log-likelihood ratio tests favour the REOP specification over the POP. However, RESET tests suggest REOP may be misspecified. This is reinforced by the discrepancy in the estimated relationship between income and health observed across the two specifications and as such (and in order to conserve space) we report only POP results here onwards[8].

The measurement model

Table 5 presents the results of POP specifications for men and women when incorporating wave 9 data on SAH. The columns of the tables represent the alternative classifications of SAH as outlined in Tables 2 and 3 above. A comparison of the SAH 5 coefficients to their respective coefficients reported in Table 4 reveals little change in the estimated magnitude of the effects of socio-economic characteristics on health, although, in general these effects increase slightly in their absolute magnitudes from zero. This is evident for both men and women. A further comparison of the coefficient estimates across the four different categorisations of SAH reveals a large degree of consistency with perhaps SAH 3 being an exception: collapsing the original five-category SAH does not appear to affect substantially the estimated relationship between socio-economic characteristics and health[9]. This finding supports the notion that the underlying latent variable model specification is appropriate. Larger differences in coefficients across the four models are observed for the estimated cut-points indicating, as one would expect, that collapsing the original five-point categorical scale into four or three categories has most effect on the measurement model that determines which category of SAH individuals report given their true latent health.

Insert Table 5 about here

A Common Index shift

Including wave 9 data has a marked effect on the estimated year dummy coefficients. The results presented in Table 4 that excluded wave 9 information showed a gradient decreasing in reported health as the cohort aged. This gradient remains after the inclusion of wave 9 data. However, the estimated coefficient for the dummy variable for wave 9 (yr9900) in general does not conform to the trend observed across the other years. This is particularly true for the SAH 5 model where the estimated coefficient for yr9900 is substantially larger than the coefficients for other years. This is not surprising given that the construction and coding of the SAH variable for wave 9 (based on the SF-36 version of SAH) is different to the construction and coding of the SAH variable at other waves. As depicted in Table 2, the coding of the variable reveals some inconsistencies across the two versions of the SAH question such that 2 represents "poor" health at waves 1-8 and 10-11 while 1 represents "poor" health at wave 9. Further, "fair" health is coded 3 at waves 1-8 and 10-11 but coded 2 at wave 9. Similarly, "good" health is coded 4 at waves 1-8 and 10-11 but coded 3 at wave 9. Only "excellent" health has common support over the coding of the two versions of the SAH question. The large negative estimated coefficient on the wave 9 dummy variable appears to reflect this shift in the mapping of "true" health to the categories of SAH as a direct artefact of the way in which the variable has been constructed.

If we consider the coefficients at wave 9 for the models adopting the SAH 4 categorisation and the approach based on the methodology of Lindley and Lorgelly (L & L) then we can see that for men the estimates are close to being consistent with the trends in the gradients observed across the other years. However, the estimates are closer to zero than true consistency with the time trends would predict. For women, the estimates do not conform to the trends observed across other years and are much closer to zero than would have been predicted on the basis of the trends. The smaller absolute value of the wave 9 coefficient would appear to reflect two things: firstly, a floor effect at wave 9 for the lowest health category of "poor" health and secondly, the larger proportion of respondents rating their health in the combined category of "good, very good" health compared to the proportion in the equivalent category of "good" health at other waves. The differences in proportions can be seen in Table 3. The combined effect of this is to induce an index shift towards reporting better health status at wave 9.

The coefficient estimate for wave 9 when considering the SAH 3 categorisation is positive for men and large and positive for women. These do not conform to the trends observed in the estimated coefficients across the other years of data. A cursory glance at Table 3 shows that the distribution of responses at wave 9 when using the SAH 3

---

[8] REOP results are available from the authors on request.

[9] The same conclusion can be drawn when considering the same analysis for all waves except wave 9 and separately for wave 9 data. These results are not presented here to conserve space and are available on request.

categorisation is skewed towards reporting better health status. Again, this is consistent with an index shift towards reporting better health at wave 9.

Reporting bias

The results reported above indicate that the data are consistent with an overall index shift at wave 9. To assess whether the observed shift in cut-points is systematically related to population sub-group characteristics we interact the wave 9 dummy variable with the socio-economic variables. Table 6 presents the results for interactions with educational attainment, income and age. Again, results are presented for POP specifications only. Models were estimated for the original five-category SAH variable together with the SAH 4 categorisation and the approach suggested by Lindley and Lorgelly. Log-likelihood ratio tests of restricted (without interactions terms) versus unrestricted (with interaction terms) models reject the restricted model for the five-category SAH variable and the Lindley and Lorgelly categorisation of SAH for women only. However, a comparison of the individual coefficients to their standard errors reveals no significant interaction terms between the socio-economic variables and the wave 9 dummy variable for any of the model specifications tested[10]. This latter result suggests that the cut-points are not systematically related to identifiable population sub-groups.

Insert Table 6 about here

Cut-point shift

In general it is not possible to distinguish between cut-point shift and heterogeneity in the latent variable equation. However if it is assumed that equations (1) and (9) are correctly specified – that is, there is no heterogeneity in the ηs and reporting bias leads to cut-point shift – then the results of the GOP can be interpreted as separating the effects of the regressors on true health from reporting bias. The former are captured by the ηs and the latter by the ι s. In order to interpret the results from the GOP model estimates, it is necessary to take into account that the top cut-point (μ^0_4) is fixed. We estimate two versions of the GOP model. In the first version the cut-points are modelled as a function of the explanatory variables corresponding to equation (7). We term this the restricted GOP model. The second version of the GOP model extends the set of explanatory variables used to test the constancy of the threshold parameters to the interactions terms between the initial set of explanatory variables used in the restricted model and the wave 9 dummy variable. We term this model the unrestricted GOP model. The unrestricted model corresponds to equations (7) and (8).

Table 7 shows GOP results for the restricted model for both men and women respondents, in particular, the coefficients obtained for the education variables, the income variable and the dummy variable corresponding to wave 9. It can be seen that for education and income variables the corresponding coefficients in the linear index model present the expected positive sign, for both men and women. The magnitudes of the ηs are similar but they tend to be smaller in the GOP models compared to the POP. When the thresholds are allowed to vary with the explanatory variables, the $\bar{Æ}^{o}_{j}$ coefficients are negative. This result implies that, as respondents increase either their level of education or their income, their assessment of health improves. Hence, individuals with higher income or level of education tend to be more optimistic about their health.

Comparing the coefficients corresponding to level of education for men, it can be seen that the absolute value of the $\bar{Æ}^{o}_{j}$ coefficients decrease with level of education, implying that individuals with higher level of education are more optimistic about their health than those with O/CSE or HND/A level. The same finding applies to the coefficients for the income variable. Similar results are found for women. The coefficients for level of education and income variables have the expected sign. However, in absolute terms, it can be seen that the coefficients are lower for women than for men. Hence, the shift in the cut-points is smaller for women.

Insert Table 7 about here

Table 8 presents the GOP results for the unrestricted model. Here all interaction terms between the set of explanatory variables and the wave 9 dummy variable are used to model both the latent health index and the cut-point thresholds. It can be seen that the coefficient estimates for the main effects of educational attainment and income are very similar to the results obtained in the restricted model. This holds for both men and women. However, the estimated coefficient on the wave 9 dummy variable differs dramatically and for men reflects closely that obtained for the POP model including interaction terms, reported in Table 6.

---

[10] This is also true for the interaction terms not presented in Table 6.

Few of the additional parameter estimates in the unrestricted GOP model achieve statistical significance at the 5% level. Of the additional (over the restricted model) 48 parameters estimated 3 achieved statistical significance for women and only 1 for men.[11] A log-likelihood ratio test of unrestricted versus restricted model fails to reject the null hypothesis that the restrictions are valid suggesting that the restricted GOP model is an adequate specification for these data.[12] Again this suggests that the change in wording of the SAH variable at wave 9 is not associated with any cut-point shift in the measurement model.

Insert Table 8 about here

6. Conclusion

Several indices and instruments are available to obtain valid information on an individual´s general level of health. The reliability of subjective information such as self-reported general health status raises concern due to it being vulnerable to several biasing factors (Knauper and Turner, 2000). However, objective measures of health status are rare in survey data and where they do exist they are often too specific to particular health conditions. Accordingly their applicability as an overall measure of an individual's health status is often limited. An appeal of measures of general health status is their ability to encapsulate and summarise a multitude of health conditions, but this is at the cost of them largely being based on subjective assessment and self-report. This is the case in the BHPS that offers a rich and diverse source of information on individual behaviour and socio-economic situation. The issue surrounding the use of self-report general health status is confounded further in the BHPS due to a change in the SAH question at wave 9. The inclusion of the SF-36 questionnaire in the ninth wave and the consequent change in the phrasing of the SAH question together with a change to the response categories made available raises challenges for the analyst wishing to conduct longitudinal work where the focus of interest is health status. It also provides an opportunity to assess the issue of reporting bias and investigate whether different groups of individuals responded in the same way to the change in measurement instrument.

This paper has explored the reliability of the measure of self-assessed health adopted in the BHPS, focusing on the rewording of the question in wave 9 and how this may affect subsequent analyses that wish to incorporate all available longitudinal information. In particular, we have investigated reporting bias due explicitly to the change in the question. For that purpose, pooled ordered probit (POP), random effects ordered probit (REOP) and generalised ordered probit (GOP) models have been used, regressing SAH on a set of relevant socio-economic variables. The results suggest that collapsing the categories of SAH to create a categorisation that has common support over the two versions of the SAH question is best achieved by collapsing to a four-category version of SAH, either by combining "very poor" and "poor" health in waves 1-8 and 10-11 and "good" and "very good" in wave 9 or by adopting the methodology proposed by Lindley and Lorgelly (2003). Collapsing the original SAH variable does not appear to affect substantially the estimated relationship between socio-economic characteristics and self-assessed health supporting the notion that the underlying latent variable model specification is appropriate. Our investigations of reporting bias when including wave 9 data suggest that the data are consistent with index shift in the measurement model, but there is little evidence that this varies with socio-economic characteristics at an individual level.

[11] For women these are: yr9900 × HND/A, yr9900 × O/CSE and yr9900 × Div/Sep as parameters for cut-point 3 (Cut 3). For men yr9900 × HND/A for cut-point 3 is significant.
[12] For men: $\theta^2_{47} \mid 42.2; p \mid 0.67$, for women: $\theta^2_{48} \mid 57.8; p \mid 0.16$.

References

Adams, P., Hurd, M.D, McFadden, D., Merrill, A., and Ribeiro, T. (2003). Healthy, wealthy and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics* 112: 3-56.

Ananth Cande V. and David G. Kleinbaum (1997). Regression models for ordinal responses: a review of methods and applications. *International Journal of Epidemiology*, Vol. 26, No. 6.

Benzeval, M., Taylor, J., and Judge, K. (2000). Evidence on the relationship between low income and poor health: Is the Government doing enough?. *Fiscal Studies*, 21(3): 375-399.

Bound, J (1990). Self-Reported versus objective measures of health in retirement models. *The Journal of Human Resources*, XXVI.

Buck N., Gershuny J., Rose D., Scott J. (1994). *Changing Hhouseholds. The BHPS 1990-1992.* ESRC Research Centre on Micro-Social Change.

Burström, B. & Fredlund, P. (2001). Self rated health: is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes?. *Journal of Epidemiology and Community Health,* 55: 836-840

Butler J., R. Moffit (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica* 50 (3): 761-764

Butler J.S., Richard V. Burkhauser, Jena M Mitchel, Theodore P. Pincus (1987). Measurement error in relf-reported health variables. *The Review of Economics and Statistics*, Volume 69, Issue 4 (Nov., 1987), 644-650.

Contoyannis, P. and Jones, A. (2003). Socio-economic status, health and lifestyle. *Journal of Health Economics*, in press.

Contoyannis P., Jones A. M., and Rice N. (2003). The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics*., in press.

Crossley T.F. and Kennedy S. (2002). The reliability of self-assessed health status. *Journal of Health Economics*, 21: 643-658

Currie, J. and Madrian, B.C. (1999). Health, health insurance and the labor market. In Ashenfelter, O. and card, D. (eds.) *Handbook of Labour Economics, Vol.3*, Elsevier, Amsterdam.

Deaton, A.S. (2003). Health, inequality and economic development. *Journal of Economic Perspectives*, XLI: 113-158.

Deaton, A..S. and Paxson, C.H. (1998). Ageing and inequality in income and health. *American Economic Review, Papers and Proceedings* 88: 248-253.

Ettner, S. (1996). New evidence on the relationship between income and health. *Journal of Health Economics*, 15: 67-85

Frechette (2001). STATA Technical Bulletin. STB-59.

Frijters, P., Haisken-DeNew, J.P. and Shields, M.A. (2003). Estimating the causal effect of income on health: evidence from post reunification East Germany, *Centre for Economic Policy Discussion Paper No. 465*, Australian National University.

Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics* 19(3), 403-420.

Grootendorst, P., Feeny, D., and Furlong, W. (1997). Does it matter whom and how you ask? Inter and intra-rater agreement in the Ontario health survey. *Journal of Clinical Epidemiology* 50: 127-136.

Idler, E. L. and Y. Benyamini (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior* 38(1): 21-37.

Idler, E.L. and Kasl, S.V. (1995). Self-ratings of health: do they also predict change in functional ability?. *Journal of Gerontology* 50B, S344-S353.

Jenkinson, C. J. et al. (1996). The U.K. SF-36: an analysis and interpretation manual. A guide to health status measurement with particular reference to the Short Form 36 Health Survey. Oxford University, Department of Public Health and Primary Care.

Kenkel, D. (1995). Should you eat breakfast? Estimates form health production functions. Health Economics, 4: 15-29.

Kerkhofs, M., Lindeboom, M. (1995). Subjective health measures and state dependent reporting errors. *Health Economics* 4, 221-235.

Knauper Barbel and Patricia A Turner (2000). Measuring health: improving the validity of health assessments. To appear in J. Dunbar-Jacob (Ed.), *Quality of Life: In the Eye of the Beholder.* University of Pittsburgh: CRCD.

Lauridsen J, T Christiansen and Unto Hakkinen (2003). Income-related inequalities in health- a sensitivity analysis using alternative weighting to categories of self-assessed health in a comparison between Denmark and Finland. Working Paper.

Lindeboom M. and van Doorslaer E. (2003). Cut-point shift and index shift in self-reported health. Working paper #2 (ECuity III). Forthcoming in Journal of Health Economics.

Lindley, J., Lorgelly, P.K. (2003). *The relative income hypothesis: does it exist over time? Evidence from the BHPS.* Paper prepared for the Health Economists' Study Group. Winter 2003, Leeds.

Long, J. Scott (1997). *Regression models for categorical and limited dependent variables.* Sage Publications.

McKelvey R.D. and Zavoina W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, Vol. 4, pp 103-120.

Murray, CJL, A Tandon, J Salomon and CD Mathers (2001). *Enhancing cross-population comparability of survey results*, GPE Discussion Paper Nr 35, WHO/EIP, Geneva.

Pudney, S. and Shields, M. (2000). Gender, race, pay and promotion in the British nursing profession: estimation of a generalized ordered probit model. *Journal of Applied Econometrics*, 15: 367 – 399.

Sadana, R, CD Mathers, AD Lopez, CJL Murray and K Iburg (2000), *Comparative analysis of more than 50 household surveys on health status*, GPE Discussion Paper No 15, EIP/GPE/EBD, World Health organisation, Geneva.

Salas, C. (2002). On the empirical association between poor health and low socioeconomic status at old age. *Health Economics*, 11: 207-220.

Smith, J.P. (1999). Healthy bodies and thick wallets: the dual relationship between health and economic status. *Journal of Economic Perspectives* 13: 145-166.

Taylor, M., Brice, J., Buck, N., and Prentice, E. (1998). *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices.* University of Essex, Colchester.

van Doorslaer, E. (1987). *Health, knowledge and the demand for medical care.* Assen, Maastricht.

van Doorslaer, E. and Gerdtham, U-G.(2003). Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data, *Social Science and Medicine*, 57: 1621-1629.

van Doorslaer, E. and Koolman, X. (2002). Explaining the differences in income-related health inequalities across European countries. ECuity II Project Working Paper #6

van Doorslaer,E. and Jones, A. M. (2003). Inequalities in self-reported health: validation of a new approach to measurement. *Journal of Health Economics*, 22(1): 61-87.

van Doorslaer, E., Jones, A.M. and Koolman, X. (2002). Explaining income-related inequalities in doctor utilisation in Europe: a decomposition approach. ECuity II Project Working Paper #5

van Doorslaer, E. et al. (1997). Income-related inequalities in health: some international comparisons. *Journal of Health Economics*, 16 (1): 93-112

van Doorslaer, E., et al. (2000). Equity in the delivery of health care in Europe and the US. *Journal of Health Economics*, 19: 553-583.

van Ourti, T. (2003). Socio-economic inequality in ill-health amongst the elderly. Should one use current income or permanent income? *Journal of Health Economics*, 22: 187-217.

Verbeek M. (2000). *Modern econometrics*. Wiley.

Wooldridge, J. (2002). *Econometric Analysis of cross section and panel data.* MIT Press.

Zeger, S. L., K.-Y. Liang and P.A. Albert (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44, 1049-60

Figure 1. Distribution of SAH for each wave

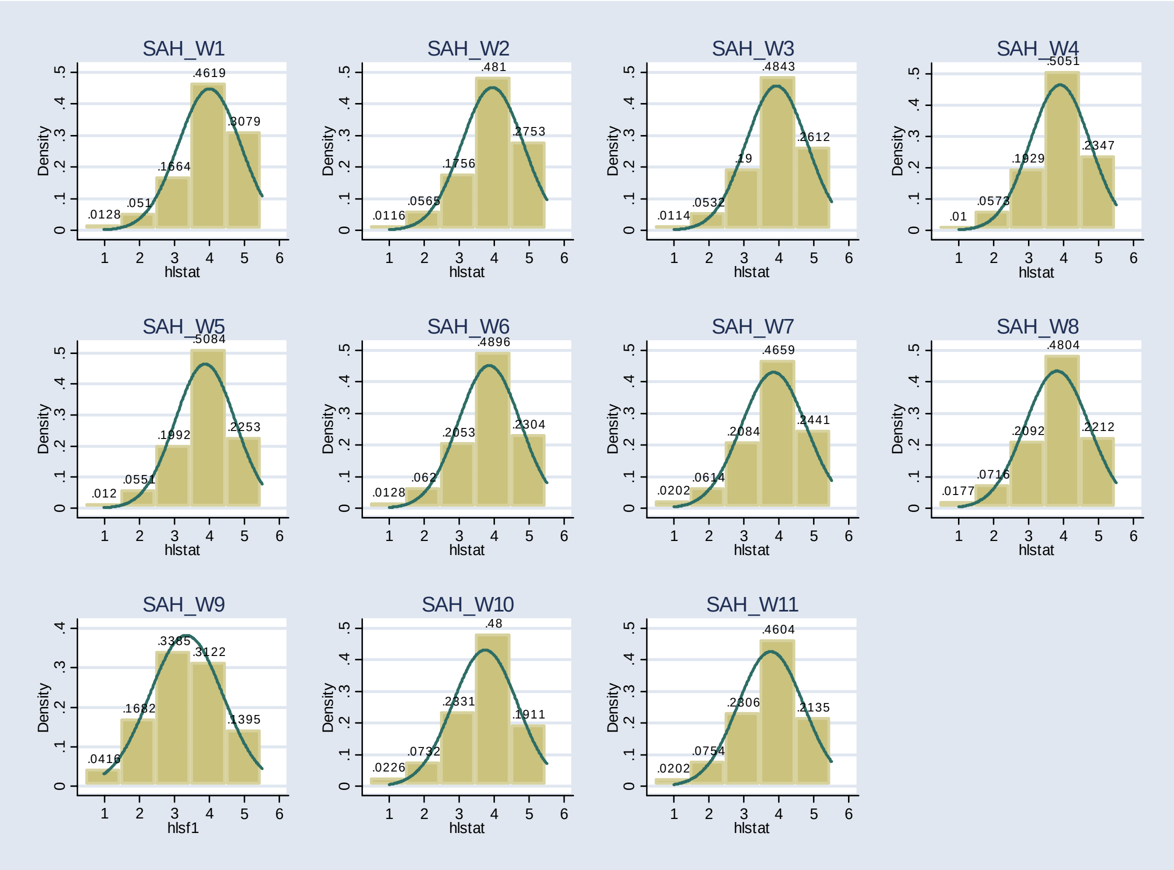

Table 1. Construction of SAH variable

Wave	Category	SAH 5 categories	SAH 4 categories	SAH 3 categories
1-8, 10-11	1	very poor	very poor, poor	very poor, poor
	2	poor	fair	fair
	3	fair	good	good, excellent
	4	good	excellent	
	5	excellent		
9	1	poor	poor	poor
	2	fair	fair	fair
	3	good	good, very good	good, very good, excellent
	4	very good	excellent	
	5	excellent		

Table 2. Frequencies for the classifications of SAH

	Wave										
	1	2	3	4	5	6	7	8	9	10	11
SAH 5											
1	1.28	1.16	1.14	1.00	1.20	1.28	2.02	1.77	4.16	2.26	2.02
2	5.10	5.65	5.32	5.73	5.51	6.20	6.14	7.16	16.8	7.32	7.54
3	16.6	17.6	19.0	19.3	19.9	20.5	20.8	20.9	33.9	23.3	23.1
4	46.2	48.1	48.4	50.5	50.8	49.0	46.6	48.0	31.2	48.0	46.0
5	30.8	27.5	26.1	23.4	22.5	23.0	24.4	22.1	14.0	19.1	21.4
SAH 4											
1	6.38	6.81	6.46	6.73	6.71	7.48	8.16	8.93	4.16	9.58	9.56
2	16.6	17.6	19.0	19.3	19.9	20.5	20.8	20.9	16.8	23.3	23.1
3	46.2	48.1	48.4	50.5	50.8	49.0	46.6	48.0	65.1	48.0	46.0
4	30.8	27.5	26.1	23.5	22.5	23.0	24.4	22.1	14.0	19.1	21.4
SAH 3											
1	6.38	6.81	6.46	6.73	6.71	7.48	8.16	8.93	4.16	9.58	9.56
2	16.6	17.6	19.0	19.3	19.9	20.5	20.8	20.9	16.8	23.3	23.1
3	77.0	75.6	74.6	74.0	73.4	72.0	71.0	70.2	79.0	67.1	67.4
L & L*											
1	6.38	6.81	6.46	6.73	6.71	7.48	8.16	8.93	7.69	9.58	9.56
2	16.6	17.6	19.0	19.3	19.9	20.5	20.8	20.9	19.8	23.3	23.1
3	46.2	48.1	48.4	50.5	50.8	49.0	46.6	48.0	48.2	48.0	46.0
4	30.8	27.5	26.1	23.5	22.5	23.0	24.4	22.1	24.3	19.1	21.4

*Following approach of Lindley and Lorgelly (2003)

Table 3. Definition of the regressors

Widow	1 if widowed, 0 otherwise
Single	1 if never married, 0 otherwise
Div/Sep	1 if divorced or separated, 0 otherwise
Non-White	1 if a member of ethnic group other than white, 0 otherwise
Degree	1 if highest academic qualification is a degree or higher degree, 0 otherwise
HND/A	1 if highest academic qualification is HND or A Level, 0 otherwise
O/CSE	1 if highest academic qualification is O level or CSE, 0 otherwise
HHSize	Number of people in household including respondent
NCH04	Number of children in household aged 0-4
NCH511	Number of children in household aged 5-11
NCH1218	Number of children in household aged 12-18
Income	Equivalised annual real household income in pounds
Age	Age in years at 1st December of current wave

Table 4. POP & REOP models excluding wave 9

	Men		Women	
	POP	REOP	POP	REOP
Degree	.377 (.058)	.536 (.080)	.381 (.052)	.338 (.077)
HND/A	.260 (.048)	.392 (.077)	.279 (.045)	.318 (.062)
O/CSE	.218 (.049)	.323 (.070)	.289 (.041)	.316 (.055)
Age	.071 (.060)	.159 (.055)	.031 (.059)	.070 (.051)
Age2	-.236 (.193)	-.508 (.176)	-.123 (.187)	-.204 (.162)
Age3	.300 (.262)	.679 (.237)	.209 (.249)	.283 (.214)
Age4	-.129 (.126)	-.328 (.114)	-.126 (.118)	-.157 (.101)
Ln(income)	.244 (.025)	.083 (.020)	.206 (.019)	.071 (.016)
yr9293	-.087 (.023)	-.079 (.035)	-.067 (.020)	-.060 (.031)
yr9394	-.133 (.024)	-.121 (.036)	-.095 (.021)	-.086 (.031)
yr9495	-.170 (.025)	-.159 (.036)	-.147 (.022)	-.137 (.031)
yr9596	-.215 (.025)	-.199 (.036)	-.160 (.023)	-.148 (.032)
yr9697	-.212 (.027)	-.193 (.037)	-.202 (.024)	-.188 (.032)
yr9798	-.237 (.028)	-.218 (.037)	-.197 (.025)	-.181 (.032)
yr9899	-.285 (.028)	-.268 (.038)	-.238 (.025)	-.222 (.033)
yr9900				
yr0001	-.368 (.029)	-.342 (.039)	-.336 (.026)	-.314 (.034)
yr0102	-.336 (.030)	-.311 (.040)	-.276 (.027)	-.260 (.034)
Cut1	.747	.175	.117	-.581
Cut2	1.514	.991	.889	.206
Cut3	2.396	1.908	1.764	1.094
Cut4	3.736	3.24	3.119	2.443
Intra-class correlation		.364		.361
Log Likelihood	-26874.1	-22339.1	-35101.4	-29513.2
RESET (p value)	.04 (.85)	15.0 (.00)	1.49 (.22)	4.23 (.04)

Notes:

1. SAH is based on the categorisation: 1: “very poor”, 2: “poor”; 3 “”fair”; 4 “good”; 5: “excellent” health.
2. Estimated coefficients for the variables Widow, Single, Div/Sep, Non-White, HHSize, NCH04, NCH511, NCH1218 have been suppressed to conserve space. Full results are available from the authors upon request.
3. Standard errors in parentheses.

Table 5. POP models including wave 9 data

	Men				Women			
	SAH 5	SAH 4	SAH 3	L & L	SAH 5	SAH 4	SAH 3	L & L
Degree	.385 (.058)	.381 (.058)	.454 (.068)	.377 (.057)	.388 (.053)	.382 (.052)	.477 (.060)	.383 (.052)
HND/A	.266 (.047)	.262 (.048)	.324 (.054)	.263 (047)	.287 (.045)	.280 (.045)	.338 (.051)	.286 (.045)
O/CSE	.222 (.048)	.219 (.049)	.268 (.055)	.219 (.048)	.292 (.041)	.289 (.041)	.326 (.045)	.290 (.040)
Age	.080 (.059)	.069 (.060)	.086 (.070)	.072 (.059)	.043 (.059)	.044 (.060)	.035 (.066)	.038 (.059)
Age2	-.266 (.192)	-.229 (.192)	-.287 (.226)	-.234 (.190)	-.163 (.186)	-.163 (.189)	-.140 (.207)	-.143 (.187)
Age3	.341 (.259)	.288 (.259)	.359 (.304)	.294 (.246)	.264 (.246)	.259 (.250)	.244 (.274)	.233 (.248)
Age4	-.148 (.125)	-.122 (.124)	-.147 (.144)	-.124 (.123)	-.152 (.116)	-.148 (.118)	-.148 (.129)	-.136 (.118)
Ln(income)	.236 (.025)	.241 (.025)	.274 (.030)	.238 (.025)	.207 (.019)	.207 (.019)	.212 (.021)	.208 (.019)
yr9293	-.085 (.022)	-.091 (.023)	-.070 (.030)	-.089 (.022)	-.067 (.020)	-.071 (.020)	-.037 (.024)	-.070 (.020)
yr9394	-.130 (.024)	-.139 (.024)	-.121 (.031)	-.136 (.024)	-.094 (.021)	-.097 (.022)	-.043 (.026)	-.096 (.022)
yr9495	-.167 (.025)	-.179 (.026)	-.136 (.033)	-.175 (.025)	-.146 (.022)	-.152 (.023)	-.057 (.027)	-.149 (.022)
yr9596	-.210 (.025)	-.222 (.026)	-.145 (.034)	-.217 (.025)	-.158 (.023)	-.164 (.024)	-.088 (.028)	-.161 (.023)
yr9697	-.207 (.027)	-.221 (.028)	-.196 (.035)	-.216 (.027)	-.200 (.023)	-.207 (.024)	-.143 (.028)	-.203 (.024)
yr9798	-.232 (.027)	-.238 (.028)	-.240 (.036)	-.233 (.027)	-.195 (.025)	-.196 (.025)	-.173 (.029)	-.193 (.025)
yr9899	-.280 (.028)	-.294 (.029)	-.278 (.036)	-.287 (.028)	-.235 (.025)	-.242 (.025)	-.194 (.029)	-.238 (.025)
yr9900	-.833 (.030)	-.288 (.027)	.183 (.038)	-.242 (.029)	-.745 (.026)	-.164 (.024)	.348 (.032)	-.128 (.026)
yr0001	-.361 (.029)	-.374 (.029)	-.356 (.035)	-.365 (.029)	-.332 (.025)	-.340 (.026)	-.274 (.030)	-.333 (.026)
yr0102	-.329 (.029)	-.344 (.030)	-.346 (.036)	-.335 (.030)	-.272 (.026)	-.280 (.027)	-.257 (.031)	-.275 (.027)
Cut1	.718	1.418	1.976	1.458	.22	1.005	1.086	.999
Cut2	1.53	2.306	2.892	2.339	1.028	1.884	1.992	1.874
Cut3	2.423	3.696		3.681	1.915	3.291		3.228
Cut4	3.733				3.245			
Log Likelihood	-30074.2	-28436.4	-16197.9	-28797.5	-39110.8	-36730.8	-23046.3	-37223.2
RESET (p value)	.29 (.59)	.00 (.98)	.03 (.87)	.02 (.88)	.17 (.68)	2.35 (.13)	3.35 (.07)	2.92 (0.09)

Notes:
1. Estimated coefficients for the variables Widow, Single, Div/Sep, Non-White, HHSize, NCH04, NCH511, NCH1218 have been suppressed to conserve space. Full results are available from the authors upon request.
2. Standard errors in parentheses.

Table 6. POP models with interaction terms, including wave 9 data

	Men			Women		
	SAH 5	SAH 4	L & L	SAH 5	SAH 4	L & L
yr9900	2.947 (1.969)	3.237 (1.525)	3.493 (1.907)	-.154 (1.677)	2.024 (1.369)	2.392 (1.737)
Interactions with yr9900:						
Degree	.111 (.068)	-.028 (.058)	-.023 (.069)	.085 (.062)	-.067 (.054)	.041 (.069)
HND/A	.092 (.051)	-.019 (.044)	.057 (.055)	.088 (.048)	-.034 (.042)	.094 (.052)
O/CSE	.071 (.052)	-.018 (.044)	.032 (.055)	.045 (.041)	-.056 (.036)	.030 (.045)
Age	-.227 (.156)	-.180 (.121)	-.221 (.150)	-.038 (.130)	-.109 (.108)	-.200 (.136)
Age2	.590 (.456)	.481 (.357)	.603 (.440)	.033 (.367)	.253 (.308)	.547 (.388)
Age3	-.665 (.566)	-.549 (.449)	-.711 (.546)	.027 (.439)	-.267 (.374)	-.645 (.468)
Age4	.270 (.254)	.225 (.204)	.301 (.245)	-.032 (.189)	.110 (.164)	.278 (.204)
Ln(income)	-.067 (.043)	-.102 (.040)	-.078 (.045)	.022 (.038)	-.050 (.032)	-.003 (.040)
Cut1	.700	1.500	1.472	.075	1.012	.991
Cut2	1.513	2.388	2.354	.885	1.891	1.866
Cut3	2.407	3.779	3.696	1.774	3.298	3.222
Cut4	3.718			3.103		
Log Likelihood	-30062.6	-28429.0	-28786.6	-39090.654	-36722.1	--37205.155
RESET (p value)	2.48 (.116)	1.89 (.169)	0.05 (.826)	13.88 (.000)	16.83 (.000)	2.95 (.086)

Notes:

1. Estimated coefficients for the variables Widow, Single, Div/Sep, Non-White, HHSize, NCH04, NCH511, NCH1218 and their interactions with the wave 9 dummy variable have been suppressed to conserve space. Full results are available from authors upon request.
2. Standard errors in parentheses.

Table 7. Generalized Ordered Probit (GOP) – restricted models

		Men			Women		
		SAH 5c (not w9)	SAH 4c (not w9)	SAH 4c	SAH 5c (not w9)	SAH 4c (not w9)	SAH 4c
Degree	Cut1	.305 (.065)	.305 (.065)	.309 (.065)	.270 (.064)	.269 (.064)	.265 (.064)
	Cut2	-.121 (.258)	-.191 (.114)	-.191 (.117)	-.221 (.133)	-.256 (.086)	-.268 (.087)
	Cut3	-.191 (.115)	-.130 (.065)	-.141 (.065)	-.257 (.086)	-.193 (.062)	-.207 (.062)
	Cut4	-.130 (.065)			-.193 (.062)		
HND/A	Cut1	.192 (.054)	.192 (.054)	.191 (.054)	.203 (.054)	.203 (.054)	.198 (.054)
	Cut2	-.079 (.121)	-.037 (.079)	-.039 (.080)	-.149 (.107)	-.115 (.068)	-.124 (.068)
	Cut3	-.037 (.079)	-.141 (.052)	-.154 (.052)	-.114 (.068)	-.128 (.052)	-.146 (.052)
	Cut4	-.141 (.052)			-.128 (.052)		
O/CSE	Cut1	.147 (.054)	.147 (.054)	.151 (.054)	.237 (.048)	.237 (.048)	.229 (.048)
	Cut2	-.105 (.117)	-.096 (.080)	-.081 (.081)	-.061 (.086)	-.069 (.062)	-.077 (.063)
	Cut3	-.096 (.081)	-.116 (.053)	-.119 (.052)	-.070 (.062)	-.085 (.046)	-.103 (.047)
	Cut4	-.116 (.053)			-.085 (.046)		
Ln(income)	Cut1	.191 (.026)	.191 (.026)	.183 (.026)	.195 (.024)	.195 (.024)	.195 (.024)
	Cut2	-.066 (.059)	-.150 (.039)	-.164 (.038)	.051 (.049)	-.001 (.031)	-.007 (.032)
	Cut3	-.149 (.039)	-.075 (.028)	-.080 (.027)	-.001 (.031)	-.028 (.024)	-.027 (.024)
	Cut4	-.075 (.028)			-.028 (.024)		
yr9900	Cut1			-.631 (.039)			-.583 (.037)
	Cut2			-.786 (.067)			-.860 (.060)
	Cut3			-.624 (.050)			-.702 (.044)
	Cut4						
Cut1		2.98 (1.55)	3.13 (1.15)	3.38 (1.15)	-4.18 (1.75)	.637 (.993)	.799 (1.01)
Cut2		3.12 (1.14)	2.84 (.830)	2.79 (.829)	.653 (.988)	2.06 (.789)	2.14 (.796)
Cut3		2.84 (.829)	2.66 (.836)	2.61 (.823)	2.072 (.787)	2.63 (.863)	2.75 (.859)
Cut4		2.65 (.838)			2.63 (.864)		
Log Likelihood		-26727.2	-26048.4	-28167.8	-34971.8	-33758.2	-36439.1

Notes:

1. Estimated coefficients for the variables Widow, Single, Div/Sep, Non-White, HHSize, NCH04, NCH511, NCH1218 have been suppressed to conserve space. Full results are available from authors upon request.
2. Standard errors in parentheses.

Table 8. Generalized Ordered Probit (GOP) - unrestricted model

		Men SAH 4c	Women SAH 4c
Degree	Cut1	.308 (.066)	.270 (.065)
	Cut2	-.183 (.115)	-.256 (.087)
	Cut3	-.127 (.065)	-.193 (.063)
HND/A	Cut1	.194 (.054)	.204 (.055)
	Cut2	-.033 (.080)	-.115 (.068)
	Cut3	-.140 (.052)	-.129 (.053)
O/CSE	Cut1	.149 (.054)	.238 (.049)
	Cut2	-.093 (.081)	-.069 (.063)
	Cut3	-.115 (.053)	-.086 (.047)
Ln(income)	Cut1	.188 (.027)	.196 (.025)
	Cut2	-.165 (039)	-.002 (032)
	Cut3	-.081 (.028)	-.029 (.024)
Yr9900	Cut1	2.365 (3.043)	.283 (2.697)
	Cut2	1.394 (6.457)	-.461 (4.878)
	Cut3	-4.010 (3.972)	-3.315 (3.561)
Interactions With yr9900:			
Degree	Cut1	.002 (.104)	-.080 (.101)
	Cut2	-.223 (.251)	-.196 (.235)
	Cut3	-.208 (.142)	-.227 (.138)
HND/A	Cut1	-.052 (.087)	-.109 (.084)
	Cut2	-.106 (.141)	-.124 (.153)
	Cut3	-.226 (.109)	-.281 (.107)
O/CSE	Cut1	.030 (.088)	-.140 (.076)
	Cut2	.208 (139)	-.120 (123)
	Cut3	-.060 (.108)	-.289 (.094)
Ln(Income)	Cut1	-.055 (.053)	.002 (.051)
	Cut2	†	-.118 (.091)
	Cut3	-.003 (.066)	.021 (.066)
Cut2		3.204 (1.153)	.639 (.994)
Cut3		2.856 (.830)	2.069 (.789)
Cut4		2.622 (.835)	2.636 (.864)
Log Likelihood		-28146.7	-36410.2

Notes:

1. Estimated coefficients for the variables Widow, Single, Div/Sep, Non-White, HHSize, NCH04, NCH511, NCH1218 and their interactions with the wave 9 dummy variable have been suppressed to conserve space. Full results are available from authors upon request.
2. Standard errors in parentheses.
3. † Maximum likelihood estimation of this model encountered numerical optimisation problems. An empirical fix to this was achieved by imposing the restriction that the coefficient on the variable Yr9900 × Ln(Income) for the second cut-point was equal to zero.