HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY of York

WP 13/31

An econometric analysis of self-assessed health: what does it mean and what is it hiding?

Nicole Au and David W. Johnston

November 2013

york.ac.uk/res/herc/hedgwp

An econometric analysis of self-assessed health: what does it mean and what is it hiding?

Nicole Au and David W. Johnston

Centre for Health Economics, Monash University, Clayton, Victoria 3800, Australia
nicole.au@monash.edu; david.johnston@monash.edu

Abstract

Self-assessed general health (SAH) is the most frequently employed health measure in economics research; however, it is poorly understood. In this paper we answer two questions: (i) what components of health does SAH measure? and (ii) does the use of SAH conceal important health effects? To answer the first question, we use a detailed health instrument and linear, logit and dynamic fixed-effects models to estimate the drivers of SAH. To answer the second question, we estimate the effects of income on disaggregated health measures using instrumental-variables fixed-effects models. We find that some health components – especially vitality – are very important to an individual when they assess their health, while others are inconsequential. We also find that this fact is partially responsible for why econometrically-sound studies find weak socioeconomic gradients in SAH. Regression results show that the effects of income on SAH are near-zero, even though income strongly affects several health components.

Keywords: General Health; Self-Assessed; Instrumental Variables; Panel Data
JEL Classification: I19

This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the author and should not be attributed to either DSS or the Melbourne Institute.

1. Introduction

Self-assessed general health (SAH), based on a simple question such as "In general, how would you rate your health?", is one of the most frequently employed health measures in economics research. It has been used to examine the relationship between health and a wide range of economic factors, including income (Ettner, 1996; Meer et al., 2003), education (Johnson, 2010; Silles, 2009), socioeconomic status (Contoyannis et al., 2004; Smith, 1999), retirement (Jones et al., 2010; McGarry, 2004) and early life experiences (Almond and Mazumder, 2005; Case and Paxson, 2008). Despite its popularity, however, SAH is a non-specific measure of health that remains poorly understood. The main advantage of using SAH is that it is probably the most feasible and inclusive measure of health status (Jylhä, 2009). It is widely accepted as a strong predictor of morbidity and mortality (see Idler and Benyamini, 1997; Jylhä, 2009 for a review). The comprehensive nature of the SAH question allows it to capture elements of health that more guided questions cannot. But, at the same time, it provides little guidance to researchers as to what individuals are thinking of when they assess their general health status. When an individual reports that their health is "poor", is it because they are in pain, tired, depressed, have limited mobility, or something else entirely? No doubt, people report poor health for a number of reasons, but given the popularity and tacit approval of SAH within economics research, it's important we better understand its structure.

A related disadvantage of using SAH is that it can encourage incorrect inference. Owing to limited space for health questions, SAH is often the only consistently collected measure of health in general population surveys, particularly across waves of longitudinal surveys. Therefore, it is often the sole measure used in many econometric analyses of the determinants and consequences of health. For this reason, important health effects may be overlooked. For example, a number of studies show that income has a small or statistically insignificant causal effect on SAH (Frijters et al., 2005; Jones and Schurer, 2011; Jones and Wildman, 2008). However, a near-zero effect on SAH may co-exist with strongly significant income effects on certain components of health (e.g. mental health or pain), especially if these elements are only weakly associated with SAH.

Given the above concerns, the objective of this paper is to answer two key questions. First, what components of health does SAH measure? And second, does the use of SAH conceal important health effects? To answer the first question we exploit rich health information contained in a nationally representative panel dataset, and use a range of fixed-effects models (linear, logit, ordered logit and dynamic) to estimate the health aspects driving SAH responses in the general population. This first analysis has parallels with influential economic studies that investigate the components driving responses to life-satisfaction and job-satisfaction questions (van Praag et al., 2003; Clark, 2001).[1] A clearer understanding of the health components driving SAH will give greater insight into research that relies on SAH, and may assist with designing interventions aimed at improving health-related quality of life. To answer the second question, we undertake an empirical demonstration that examines the effect of income on health using an instrumental-variable fixed-effects (IV-FE) methodology. We compare the estimates obtained when health is measured using SAH, to estimates obtained when using disaggregated measures of health.

To the best of our knowledge, there exists very little research that directly answers the above two questions.[2] A number of papers from different economic literatures contain regression models of SAH on reported health conditions; however, their motivations are quite different to ours. For example, Powdthavee and van den Berg (2011) use the British Household Panel Survey to estimate monetary values for a number of long-term health conditions using several well-being measures, including SAH. They find that people tend to focus more on physical conditions than on mental conditions when evaluating their subjective health status. An important feature of our methodology that differs from this paper is the use of a detailed health instrument, which allows us to capture a broader range of health dimensions and to understand the SAH responses of the general population; the majority of whom do not have long-term health conditions.

Our findings indicate that when an individual assesses their general health the most important consideration is whether they are feeling full of life and energetic or worn out and tired (vitality). Other important considerations are whether ill health is limiting mobility (physical functioning) and causing pain (bodily pain). These results are exceptionally robust to a number of different specifications and estimators, and are consistent across subgroups stratified by gender, age, education, and experience with ill-health. They also hold when we

[1] van Praag et al. (2003) find that the three main determinants of overall life satisfaction are satisfaction with finances, health, and work. Clark (2001) finds that what matters most in a job is satisfaction with pay and security, followed by the use of initiative, the work itself and hours of work.

[2] Several studies in the health and psychology literature have investigated health factors that are associated with SAH using cross-sectional data (see for example, Andersen and Lobel, 1995; Benyamini et al., 2000; Singh-Manoux et al., 2006). However, it is difficult to draw any strong conclusions from these studies, primarily due to differences in scales used to measure health dimensions within a single study, and the high variability in the choice of included health dimensions.

condition out threshold-specific unobserved heterogeneity by estimating a conditional fixed-effects logit model for each threshold value (Jones and Schurer, 2011).

Estimates from IV-FE models demonstrate that although income has a near-zero statistically insignificant effect on SAH, it has a significantly positive effect on certain components of health – particularly health components related to the inability to carry-out day-to-day activities due to ill-health. Importantly, these particular health components are weakly associated with SAH responses and so may explain why econometrically-sound studies find weak socioeconomic gradients in SAH. These findings imply that particular care must be taken when interpreting econometric models of SAH, because a zero SAH effect does not mean zero effects in all health components or even in the majority of health components. We highlight the need for large, nationally representative surveys to more frequently include disaggregated health measures, and when available, for analyses to include estimates using disaggregated health measures alongside estimates using SAH.

2. Data, Definitions and Descriptive Statistics

2.1. The Household, Income and Labour Dynamics in Australia (HILDA) survey

We use data from the HILDA survey, a continuing nationally representative longitudinal survey of Australian households that began in 2001. The main motivation for using HILDA is that it is the only nationally representative panel survey that includes quality annual information on socioeconomic characteristics as well as a detailed generic health survey.[3] Demographic and socioeconomic data are collected through face-to-face interviews, while most information on health and lifestyle behaviours is collected through a self-completion questionnaire. In this paper we use all 11 currently available waves of HILDA (2001 to 2011). After omitting respondents with missing information on the outcome variables or covariates, and respondents who only appear in one wave (due to the exclusive use of fixed-effects models), the main estimation sample includes 104,143 observations on 16,799 respondents aged 18 to 80.

2.2. Self-assessed health

The first part of the self-complete questionnaire in HILDA is the SF-36, a generic health survey comprising 36 questions. The main outcome variable in this study is based on the first

[3] The British Household Panel Survey contains the SF-36 survey, but only in waves 9 and 14 (1999 and 2004).

question of the SF-36, which states “In general, would you say your health is: excellent, very good, good, fair or poor”. The responses are coded as: 1 = Poor (3% of all observations), 2 = Fair (13%), 3 = Good (36%), 4 = Very Good (37%), and 5 = Excellent (11%).

Looking at year-to-year changes in self-assessed health (SAH), 60% report no change, 17% and 1% report a one and two unit improvement respectively, and 19% and 2% report a one and two unit worsening respectively. Movements of greater than two units are rare (0.3%). The most common changes in SAH occur between good and very good health. Interestingly, the vast majority of these movements are not associated with reported health conditions. For example, of the 18,221 observed deteriorations in SAH, only 28% correspond to individuals with a health condition. The similar figure for improvements in SAH equals 23%. This suggests that the incidence of serious illness is unlikely to be the driver of SAH for most individuals.

2.3. SF-36 health dimensions

The SF-36 is widely used to measure overall health-related wellbeing in general and specific populations (Ware, 2000). It has been psychometrically evaluated and extensively tested for its reliability and validity in many countries (Ware, 2000). The SF-36 yields summary measures for eight health dimensions: 1) general health; 2) vitality; 3) physical functioning; 4) bodily pain; 5) mental health; 6) social functioning; 7) role limitations due to physical health; and, 8) role limitations due to emotional problems. These eight health dimensions were selected from 40 included in the Medical Outcomes Study and represent the most frequently measured concepts in widely-used health surveys (for more information see Tarlov et al., 1989). Although the SF-36 does not include all possible health dimensions, it has been shown to correlate highly with most other common general health concepts that are not included in the SF-36, suggesting that the eight dimensions do well to capture most health dimensions of interest (Ware, 2000).[4]

For our analysis we include all seven non-general health dimension scores as the key explanatory variables (see Table 1 for a description of each dimension). We omit the general health dimension because it contains the SAH question. All dimensions are scaled from 0 to 1, where 0 represents the worst (unhealthiest) score and 1 represents the best (healthiest) score. We recognise that dimensions which have more items (questions) may be more sensitive than

[4] One notable omission is sexual functioning, which has been shown to correlate relatively weakly with the SF-36 (Ware, 2000). Other omissions, such as sleep adequacy, cognitive functioning, eating, self-esteem, and communication, are partly represented by one or more of the eight included dimensions.

those which include fewer items, and that differences in the sample variations of the dimension scores may compromise the comparability of effect sizes. We therefore estimate robustness analyses using alternative measures for the health dimensions (see Section 3.2) and find that the main results are robust.

Table 1: Description of Health Variables and Covariates

Label	Description	Mean	Std
SAH	1 item ranging from poor to excellent (1-5)	3.410	0.950
Health Explanatory Variables			
Vitality	4 item index measuring vitality (0-1)	0.605	0.195
Physical functioning	10 item index measuring physical functioning (0-1)	0.848	0.215
Bodily pain	2 item index measuring bodily pain (0-1)	0.740	0.239
Mental health	5 item index measuring mental health (0-1)	0.746	0.169
Role physical	4 item index measuring problems carrying out work or other activities due to physical health (0-1)	0.803	0.350
Social functioning	2 item index measuring disruptions to social activities due to physical or mental health (0-1)	0.833	0.228
Role emotional	3 item index measuring problems carrying out work or other activities due to mental health (0-1)	0.842	0.318
Future health	1 item measuring expectations of future health (0-1)	0.647	0.283
Easily sick	1 item measuring ease of getting sick (0-1)	0.804	0.250
Health condition	Long-term health condition, impairment or disability that restricts activities (dv)	0.246	0.431
Healthy lifestyle	Scaled count of good health behaviours: exercise ≥ 3 times/week; don't smoke; < 5 standard drinks on any one occasion (0-1)	0.698	0.269
Covariates			
Age	Age (dv)	44.27	15.96
Age squared	Age squared divided by 100 (dv)	22.15	15.02
Age cubed	Age cubed divided by 10000 (dv)	12.17	11.77
Diploma / certificate	Highest qualification is diploma or certificate (dv)	0.318	0.466
University	Highest qualification is university degree (dv)	0.233	0.423
Married /	Married or cohabitating (dv)	0.685	0.465
Divorced / separated	Divorced or separated (dv)	0.127	0.333
Number of children	Number of children in the household (dv)	0.619	1.017
Employed	Employed full-time or part-time (dv)	0.687	0.464
Unemployed	Unemployed (dv)	0.029	0.168

Note: Sample statistics calculated using pooled sample of 104143 observations. The abbreviation dv denotes dummy variable. The figures in parentheses in the 'Description' column denote the ranges of the non-binary variables.

A correlation matrix of SAH and the seven health dimensions is shown in Table 2. The lower triangle, which presents correlations using the raw data, shows that SAH is most strongly correlated with vitality (0.54), followed closely by bodily pain (0.53) then physical functioning (0.52). It is weakly correlated with emotional role limitations (0.35). The strongest correlation in the matrix is between mental health and vitality (0.69). The upper triangle shows correlations between annual changes in the variables, i.e. $corr\left(sah_t - sah_{t-1}, h_{jt} - h_{jt-1}\right)$ where h_j is health dimension j. Although correlations are smaller overall, they lead to broadly similar conclusions. One notable difference, however, is that physical functioning (0.16) drops behind mental health (0.18), social functioning (0.21) and physical role limitations (0.19) in terms of strength of correlation with SAH.

Table 2: Correlations of Levels and Changes in Self-Assessed Health and SF-36 Health Dimensions

	SAH	VT	PF	BP	MH	SF	RP	RE
Self-assessed health – SAH		0.26	0.16	0.20	0.18	0.21	0.19	0.12
Vitality – VT	0.54		0.18	0.26	0.52	0.42	0.27	0.28
Physical functioning – PF	0.52	0.40		0.24	0.10	0.19	0.26	0.10
Bodily pain – BP	0.53	0.50	0.57		0.16	0.34	0.41	0.12
Mental health – MH	0.40	0.69	0.26	0.36		0.45	0.14	0.37
Social functioning – SF	0.49	0.61	0.48	0.57	0.63		0.39	0.39
Role physical – RP	0.49	0.47	0.59	0.64	0.33	0.60		0.22
Role emotional – RE	0.35	0.47	0.33	0.36	0.56	0.62	0.47	

Notes: Figures are estimated correlations of the variables in levels (lower triangle) and of annual changes in the variables (upper triangle). Sample size is 104143.

2.4. Other health variables

Most components of health will influence SAH via one (or more) of the included SF-36 health dimensions, and for this reason our primary specification focuses on the seven SF-36 variables. However, it is possible that some health components are not captured by the SF-36 health dimensions, and so we include additional health variables in supplementary specifications. To account for the potential influence of medical diagnoses (Jylhä, 2009; Krause and Jay, 1994), we add an indicator for a reported long-term health condition, impairment or disability. Qualitative studies have shown that some people take health behaviours into account when evaluating their health (Krause and Jay, 1994; Manderbacka, 1998). We therefore add a scaled index (from 0 to 1) that counts the number of healthy behaviours, as measured by physical activity, smoking status and alcohol consumption.[5] We also recognise that expectations about future health and perceptions about own vulnerability to illness may influence one's perceived health (Andersen and Lobel, 1995; Jylhä, 2009). To account for these aspects, we include the two items from the SF-36 that reflect future health expectations and how easily the respondent feels they get sick. These items were omitted from the primary model because they fall under the 'general health' SF-36 dimension along with SAH. Finally, we add lagged SAH, since it is possible that current SAH is affected by past health shocks, even if the individual is no longer ill. This may occur if individuals make relative comparisons with past health states (Foster et al., 2012).

3. What does Self-Assessed Health Measure?

3.1. Methodological Approach

We conceptualise responses to the general health question as the end product of a three-stage process.[6] First, an individual *i* in period *t* considers relevant health components, such as their functional status, physical sensations, symptoms, medical diagnoses, and genetic dispositions $(h^*_{1it}, h^*_{2it}, \dots, h^*_{Jit})$. Second, these latent health components are combined and transformed into a latent assessment of overall health (sah^*_{it}) through a function $f_1(\cdot)$ that varies across individuals according to reference groups, health expectations and earlier health experiences (denoted by R_{it}). Finally, as people select the response option that best describes their general health, latent SAH is transformed on to the ordinal 1 (poor) to 5 (excellent) scale (sah_{it}) through another function $f_2(\cdot)$, which also varies by individual-level characteristics, such as culture, personality and language (denoted by C_{it}). This evaluation process can be described by the equations:

$$sah^*_{it} = f_1\left(h^*_{1it}, h^*_{2it}, \dots, h^*_{Jit}; R_{it}\right) \quad (1a)$$

$$sah_{it} = f_2(sah^*_{it}; C_{it}) \quad (1b)$$

$$\Rightarrow sah_{it} = f_3\left(h^*_{1it}, h^*_{2it}, \dots, h^*_{Jit}; R_{it}, C_{it}\right) \quad (1c)$$

[5] A score of 1 indicates the individual is 'healthy' in all three lifestyle domains, while 0 indicates 'unhealthy' in all three domains. Fruit and vegetable consumption are other commonly measured lifestyle factors but are not included in the index because they are only available in two waves of HILDA.
[6] This conceptualisation is based on the framework describing the individual health evaluation process in Figure 1 from Jylhä (2009), which integrates information from social and biological disciplines.

These equations make clear that SAH is not only dependent upon health components, but also upon non-health factors that influence how an individual views good health and how health is translated on to an ordinal scale (i.e. vectors R_{it} and C_{it}). These non-health factors can be collectively thought of as response heterogeneity. Naturally, we recognise self-assessments are not necessarily based on logical steps of cognitive reasoning such as these, but the model provides a useful conceptual framework for our empirical investigations.

Given this framework, our methodological approach is to approximate equation (1c) with regression models of SAH that include as many of the relevant health components as our data allows (i.e., SF-36 health dimensions and additional health factors described in Section 2.4). To control for response heterogeneity, we include a vector of time-varying characteristics: age, educational attainment, employment status, marital status, and number of children (see Table 1). We additionally control for individual-level fixed-effects, which represent time-invariant characteristics that may influence reporting behaviour (e.g. culture, personality and language). The fixed-effects may also reflect time-invariant health components, such as known genetic dispositions and chronic health conditions that span the panel. The error term in our regression model will capture time-varying health components that are poorly reflected by the SF-36 health dimensions and other included health variables. It will also capture time-varying response heterogeneity not controlled for by the observed characteristics, and random noise due to current mood and immediate contextual information.

We employ several alternative estimators, including ordinary least squares (OLS), linear fixed-effects (FE), conditional ordered logit FE, and linear dynamic FE. Our workhorse model is linear FE regression:

$$sah_{it} = \alpha_{1i} + H'_{it}\beta_1 + X'_{it}\gamma_1 + \varepsilon_{1it} \tag{2}$$

where α_{1i} is the individual-level fixed-effect, H_{it} is a vector of observed health components, X_{it} is a vector of observed time-varying characteristics, and ε_{1it} is an error term. A limitation of model (1) is its disregard for the ordinal nature of SAH. To overcome this limitation we also estimate ordered-logit FE models:

$$\begin{aligned} &sah^*_{it} = \alpha_{2i} + H'_{it}\beta_2 + X'_{it}\gamma_2 + \varepsilon_{2it} \\ &sah_{it} = k \iff sah^*_{it} \in [\tau_k, \tau_{k+1}] \end{aligned} \tag{3}$$

where sah^*_{it} is a latent SAH index, τ_k are SAH thresholds increasing in k, and ε_{2it} is an idiosyncratic logit-distributed error term. The estimation method for this model, developed by Ferrer-i-Carbonell and Frijters (2004), involves using an individual-specific threshold to collapse the ordered outcome variable (SAH) into a binary variable, and then applying Chamberlain's (1980) conditional approach for estimating fixed-effects logit models. A commonly used approximation to this method involves using the within-individual mean values of SAH as the individual-specific threshold (Jones and Schurer, 2011). It is important to note that using a fixed-effects ordered-logit model has its drawbacks. There is a large reduction in observations due to the omission of individuals without time-variation in SAH, and marginal effects cannot be calculated without additional untestable assumptions. It is for these reasons, coupled with the qualitative similarity of the linear and ordered results, that we primarily present estimates from linear models.

In addition to linear FE and ordered-logit FE models, we estimate linear dynamic FE models. These models are used in order to evaluate the importance of past SAH levels on current SAH. The dynamic FE model is specified as:

$$sah_{it} = \alpha_{3i} + \delta sah_{it-1} + H'_{it}\beta_3 + X'_{it}\gamma_3 + \varepsilon_{3it} \tag{4}$$

and is estimated using a system GMM specification that contains the SAH equation in levels and differences (Blundell and Bond, 1998). In this method, lagged first differences are used as instruments for the equation in levels and lagged levels are used as instruments for the equation in first differences. The main assumptions underlying this method are no autocorrelation in the idiosyncratic error ε_{3it}, and no correlation between the fixed effects and the first difference of the first observation of SAH.

In the above models, no explicit allowance is made for heterogeneity in the threshold levels used to assess health. For example, in model (3) it is assumed that for all i and k the thresholds $\tau_k = \tau_{ik}$ (no cut-point shifts). This is a potential limitation given a number of studies show that some individuals with the same levels of true health report their SAH differently (see for example, Bago d'Uva et al., 2008; Baron-Epel et al., 2005; Bound, 1991; Etilé and Milcent, 2006; Groot, 2000; Jürges, 2007; Kerkhofs and Lindeboom, 1995; Lindeboom and Van Doorslaer, 2004; Shmueli, 2003). We explore this issue by re-estimating our models for subgroups defined by characteristics shown to determine reporting heterogeneity; namely, gender, age, education, and history of ill health. We also follow the approach in Jones and Schurer (2011) and estimate a series of conditional fixed-effects logit models, which conditions out threshold-specific individual heterogeneity.

3.2. Main Results

Estimated associations between SAH and the seven SF-36 health dimensions from OLS, linear FE and ordered-logit FE models are presented in Columns (1) to (3) of Table 3. The most important finding from this set of estimates is that vitality is by far the strongest predictor of SAH. The vitality coefficient of 1.406 in Column (1) indicates that moving from lowest vitality (a score of 0) to highest vitality (a score of 1) increases SAH by 1.4 units, holding other health dimensions constant. This coefficient is significantly different to all other health dimension coefficients at the 1% level. The next strongest predictors are physical functioning and bodily pain, which have roughly half the effect of vitality in each of the models. The remaining significant predictors of SAH are mental health (roughly one-quarter of the vitality effect), role physical (roughly one-sixth of the vitality effect), and social functioning (roughly one-seventh of the vitality effect). Role emotional is an insignificant predictor of SAH in the linear and ordered logit fixed-effect models.[7]

The reported R-squared values in Table 3 equal the squared correlation of SAH and predicted SAH using the estimated coefficients (but not estimates of the fixed-effects). The values suggest that within-individual variation in the health dimensions have considerable predictive power, but that there is also a significant proportion of the variation that is unexplained. The R-squared associated with model (2) increases to 0.75 if we also include the predictive power of the fixed-effects, suggesting that time-invariant factors (e.g. response heterogeneity) are also important. A possible contributor to the remaining variation is random noise due to current mood and immediate contextual information, such as day of the week effects (Taylor, 2006) and interviewer effects (Bateman and Mawby, 2004).

The relative size of the health dimension coefficients could be sensitive to how the variables are standardised, given that the dimensions have different variances and are constructed from different numbers of items. In Table 3 the variables are scaled from 0 (worst) to 1 (best). An alternative approach is to standardise the variables such that they have a mean of zero and a standard deviation of one. Another approach is to use only one item from each dimension. Estimates associated with these two approaches show that the Table 3 results are robust to alternative measurement scales (results available upon request). A fourth approach

[7] As is commonly found, education, marriage, and employment are positively associated with SAH, and age is negatively associated with SAH in the OLS model. Full results are available upon request.

to evaluating the strength of association is to use partial R-squared values. For model (1), the partial R-squared values for vitality, physical functioning, bodily pain, mental health, role physical, social functioning and role emotional equal 0.060, 0.028, 0.019, 0.002, 0.004, 0.001 and <0.001, respectively. Therefore, this alternative measure gives results, in terms of the ordering and relative strength of dimensions, that are broadly consistent with Table 3.[8]

Table 3: Regression Models of Self-Assessed Health

	Alternative Estimators			Alternative Covariate Sets		
	OLS (1)	FE (2)	Ordered Logit FE (3)	FE (4)	FE (5)	Dynamic FE (6)
Vitality	1.406**	0.914**	3.439**	-	0.787**	0.709**
	(0.028)	(0.022)	(0.089)		(0.021)	(0.030)
Physical functioning	0.794**	0.471**	1.716**	0.538**	0.392**	0.367**
	(0.024)	(0.021)	(0.077)	(0.021)	(0.020)	(0.029)
Bodily pain	0.591**	0.398**	1.641**	0.467**	0.334**	0.295**
	(0.021)	(0.015)	(0.063)	(0.015)	(0.015)	(0.021)
Mental health	0.327**	0.236**	0.893**	0.679**	0.154**	0.163**
	(0.031)	(0.024)	(0.101)	(0.023)	(0.024)	(0.034)
Role physical	0.201**	0.165**	0.561**	0.207**	0.137**	0.125**
	(0.012)	(0.009)	(0.040)	(0.010)	(0.009)	(0.013)
Social functioning	0.179**	0.159**	0.431**	0.259**	0.143**	0.150**
	(0.021)	(0.017)	(0.069)	(0.017)	(0.016)	(0.023)
Role emotional	-0.045**	0.016	-0.005	0.026**	0.012	0.020
	(0.012)	(0.009)	(0.040)	(0.009)	(0.009)	(0.013)
Future health	-	-	-	-	0.308**	0.236**
					(0.011)	(0.015)
Easily sick	-	-	-	-	0.316**	0.216**
					(0.013)	(0.018)
Health condition	-	-	-	-	-0.141**	-0.096**
					(0.007)	(0.009)
Healthy lifestyle	-	-	-	-	0.164**	0.137**
					(0.011)	(0.016)
Lagged SAH	-	-	-	-	-	0.098**
						(0.007)
R-squared	0.449	0.393	-	0.346	0.470	0.527
Individuals	16799	16799	12091	16799	16799	14463
Observations	104143	104143	89228	104143	104143	76746

Notes: Figures are estimated coefficients. Standard errors clustered at the individual-level are shown in parentheses. * and ** denote significance at .05 and .01 levels. The outcome ranges from 1 (poor) to 5 (excellent). The covariates are scaled from 0 (worst) to 1 (best), except 'Health condition' which is a binary indicator of a long-term health condition, and lagged SAH which ranges from 1 (poor) to 5 (excellent). Included in each model but not shown are the set of time-varying covariates shown in Table 1.

[8] Another alternative approach is to include all of the individual SF-36 items as covariates. With this approach, the largest estimated coefficient is for the item "Did you have a lot of energy?", followed by "Did you feel full of life?" and "How much bodily pain have you had during the past 4 weeks ?".

The results are also robust to the inclusion of higher-order polynomial functions of each health dimension. If cubic functions of each dimension are included, the partial R-squared values for vitality, physical functioning, bodily pain, mental health, role physical, social functioning and role emotional equal 0.053, 0.046, 0.014, 0.002, 0.004, 0.002 and 0.001, respectively. These results show that in this expanded specification, vitality is again the most important health dimension, followed by physical functioning.[9]

An unexpected result in Columns (1) to (3) is the relatively weak association between the mental health dimension and SAH. A possible explanation for this finding is that mental health conditions, such as depression and anxiety, largely impact upon SAH through loss of vitality rather than through feelings of melancholy and nervousness. We explore this possibility by re-estimating the linear fixed-effect model without vitality. The results in Column (4) show that the mental health dimension now has the largest association with SAH, validating this explanation. More generally, it is interesting to determine whether mental health or physical health is a greater predictor of SAH. To do so, we regressed SAH against the first two predicted factors from a principal components analysis of the seven health dimensions.[10] The estimated fixed-effects coefficients equalled 1.820 and 1.911 for the mental health and physical health factors, respectively, suggesting that physical health explains a slightly greater proportion of SAH. This finding is robust to the comparison of mental and physical health partial R-squared values.

Columns (5) and (6) in Table 3 present estimates from models with additional health variables. First, variables representing future health expectations, perceptions about own vulnerability to illness, long term health condition or disability, and healthy behaviours are added (see Section 2.4 for definitions). Each of these variables is significantly associated with SAH, and their inclusion diminishes the estimated coefficients of the SF-36 health dimensions. Of the four variables, ‘future health’ and ‘easily sick’ have the largest coefficients, with a movement from the lowest to highest values increasing SAH by around 0.3 units. Importantly, the inclusion of these variables has not altered the order of importance of the SF-36 health dimensions, and the vitality coefficient remains comparatively large. Next, lagged SAH is added to the model, which necessitates the use of the Blundell and Bond (1998) GMM system estimator (see Section 3.1 for details). The results in Column (6) show

[9] The additional 14 variables (7 squared terms and 7 cubic terms) do not substantially increase the overall fit of the model: the OLS and FE R-squared values increase from 0.449 to 0.461 and from 0.393 to 0.400, respectively. This suggests that the linear functions used in Table 3 are a reasonable approximation.

[10] The first factor loaded heavily on mental health (0.908), vitality (0.760) and role emotional (0.733), while the second factor loaded heavily on physical functioning (0.844), role physical (0.816) and bodily pain (0.803). Each factor was standardised to range from 0 to 1.

that even conditional on the 11 contemporaneous health variables, SAH last year is a significant predictor of SAH today: moving from poor (1) to excellent (5) health increases SAH by 0.392 units (4 x 0.098), which is roughly equivalent in size to the effect of moving from lowest to highest physical functioning. The lagged SAH effect could be driven by reference point effects, persistent health shock effects not adequately captured by the contemporaneous health variables, or time-varying response heterogeneity.

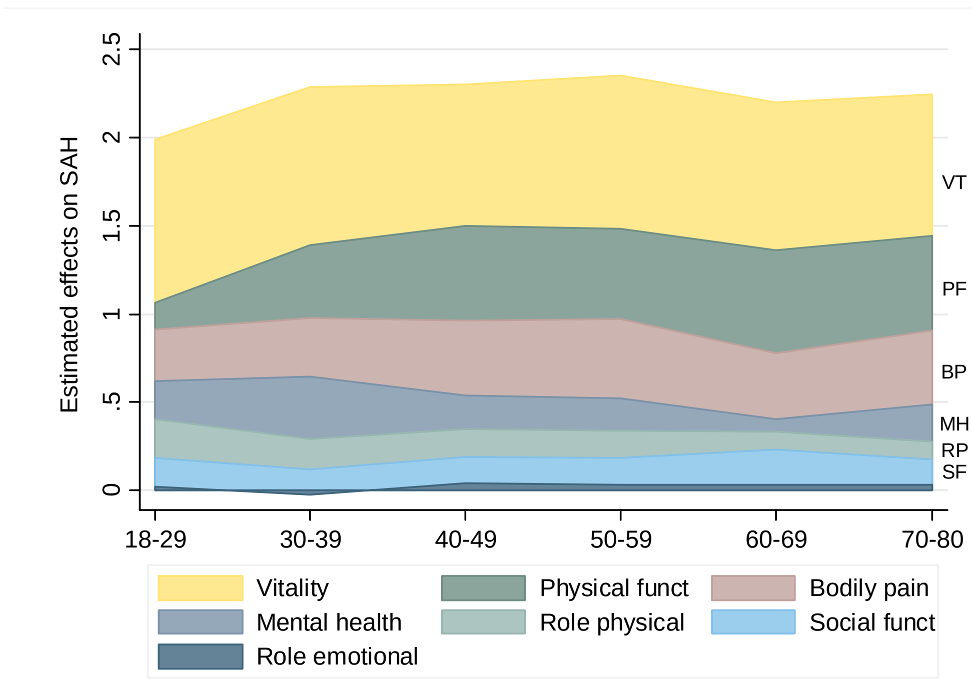
In summary, the results in Table 3 demonstrate that when an individual assesses their health, a major consideration is whether they are feeling full of life and energetic or worn out and tired (vitality). The psychology literature suggests that vitality is associated with both psychological and physical factors that impact on the energy available to one's self (Ryan and Frederick, 1997). As our models control for a range of other mental and physical health dimensions, and serious long term health conditions, the persistently strong influence of vitality on SAH suggests that vitality is perhaps partially driven by minor short-term ailments, such as cold and flu symptoms, headaches, and lack of sleep (Ryan and Frederick, 1997). Table 3 also shows that other important considerations in assessing SAH are whether ill health is limiting mobility (physical functioning) and causing pain (bodily pain). Interestingly, having a serious long-term health condition is a relatively weak independent predictor of SAH, which suggests health conditions are generally well captured by the SF-36 dimensions.

3.3. Sub-group effects

The above results provide associations between health dimensions and SAH for an average individual. It is possible that these estimates mask considerable heterogeneity. For example, if individuals assess their health relative to their same-aged peers, then it is possible that the associations will vary across the age distribution. Also, ill-health in certain dimensions will be common at some ages and rare at other ages. Other likely determinants of reporting heterogeneity are gender, education, and history of ill health. In this subsection, we re-estimate linear FE regressions by sub-groups in order to ascertain the extent of heterogeneity.

Figure 1 presents estimated associations for the seven SF-36 dimensions by six age groups (18-29, 30-39, 40-49, 50-59, 60-69 and 70-80). The highest point of the graph at each age group represents the total change in SAH from moving from worst health in each dimension to best health in each dimension. The total change is lowest for the youngest age group (equals 1.99) and is highest for the 50-59 age group (equals 2.35), but overall the variation with age is small. The estimates for vitality are also reasonably homogenous: the smallest estimate equals 0.80 (age 70-80) and the largest estimate equals 0.93 (age 18-29). At every age vitality has the largest estimated association with SAH. The most heterogeneous estimates are for physical functioning. They range from 0.15 (age 18-29) to 0.58 (age 60-69), suggesting that physical functioning is more important at older ages than at younger ages. This heterogeneity is likely due to the strong negative relationship between physical functioning and age: the mean score equals 0.93 at age 18-29 and equals 0.63 at age 70-80. By comparison, there is no relationship between vitality and age – people can feel worn out and tired at any age.

Figure 1: Stacked Regression Coefficient Estimates from Linear Fixed-Effect Regression Models by Age Group



Estimated associations by gender, education, and history of ill health are presented in Columns (1) to (6) in Table 4. Overall, the estimates do not differ greatly across gender or education levels.[11] The estimates do differ somewhat by history of health, which is defined using mean SAH in waves 1 to 3. A history of poor health corresponds to a mean SAH value

[11] Consistent with all other models, partial R-squared values provide the same ordering and relative strength of dimensions as the estimated coefficients do.

< 3 (19% of respondents) and a history of excellent health corresponds to a mean SAH value > 4 (20% of respondents). For this subgroup analysis, only waves 4 to 11 are used in the regressions. One notable difference across these two groups is that mental health is a much more important determinant of SAH for the excellent health group (more than twice the size of the association compared to the poor health group). Another point of difference across these two groups is in the estimated R-squared. For the poor health group the R-squared equals 0.324, while for the excellent health group the R-squared equals 0.130. This result is driven by the much higher variances in the SF-36 health dimensions of the poor health group compared with the excellent health group, which report consistently high health.[12]

Table 4: Fixed-Effect Regression Models of Self-Assessed Health by Subgroups

	Female (1)	Male (2)	Low/Med Education Level (3)	High Education Level (4)	History of poor health (5)	History of excellent health (6)
Vitality	0.873**	0.962**	0.882**	0.958**	0.797**	0.862**
	(0.029)	(0.032)	(0.027)	(0.038)	(0.057)	(0.069)
Physical funct.	0.533**	0.408**	0.439**	0.580**	0.583**	0.403**
	(0.030)	(0.029)	(0.024)	(0.043)	(0.054)	(0.085)
Bodily pain	0.385**	0.413**	0.403**	0.393**	0.351**	0.471**
	(0.021)	(0.022)	(0.018)	(0.027)	(0.044)	(0.049)
Mental health	0.255**	0.209**	0.246**	0.212**	0.162**	0.337**
	(0.033)	(0.036)	(0.029)	(0.045)	(0.059)	(0.087)
Role physical	0.166**	0.164**	0.166**	0.159**	0.153**	0.207**
	(0.013)	(0.014)	(0.011)	(0.016)	(0.021)	(0.038)
Social funct.	0.170**	0.146**	0.149**	0.179**	0.180**	0.235**
	(0.023)	(0.025)	(0.020)	(0.030)	(0.039)	(0.072)
Role emotional	0.013	0.020	0.010	0.041*	0.016	0.062
	(0.012)	(0.014)	(0.011)	(0.017)	(0.020)	(0.043)
R-squared	0.407	0.382	0.397	0.358	0.324	0.130
Individuals	8679	8121	12490	5134	2276	2217
Observations	54751	49392	70029	34114	11103	11582

Notes: Figures are estimated coefficients. Standard errors clustered at the individual-level are shown in parentheses. * and ** denote significance at .05 and .01 levels. The dependent variable in all models ranges from 1 (poor) to 5 (excellent). All health covariates are scaled from 0 (worst) to 1 (best). Included in each model but not shown are the set of time-varying covariates shown in Table 1. Low/Med education is defined as having no post-school qualifications or having a vocational certificate. High education is defined as having at least a diploma or bachelor degree. History of health is based on average SAH in the first 3 waves, which are not included in the model. Poor health is an average score <3 and excellent health is an average score >4.

[12] We also tested whether the effects noticeably differed for those who speak a language other than English at home (13% of respondents). The estimated coefficients and R-squared values were similar to those for native English speakers.

3.4. Alternative outcome measures

The ordered SAH variable used in Tables 3 and 4 is a frequently used measure of general health in economics studies. However, researchers also frequently dichotomise this variable to denote poor health or excellent health. In this section we investigate whether the results from Section 3.2 are robust to binary transformations of ordered SAH. Using a systematic approach, we estimate a conditional FE logit model for each of the four binary variables into which SAH can be dichotomised (SAH>1, SAH>2, SAH>3 and SAH>4). For instance, this means the first dependent variable (SAH>1) equals one if SAH is 'fair' or better, and equals zero if SAH equals 'poor'. This approach is equivalent to the methodology used in Jones and Schurer (2011) to examine whether there is heterogeneity (nonlinearity) in the effects of income across values of the latent SAH index. Consequently, we apply the interpretation used by Jones and Schurer (2011) to our results. For example, a positive coefficient for vitality in the SAH>1 model and a zero coefficient for vitality in the SAH>4 model implies that an improvement in vitality improves general health at the lower end of the health distribution but not at the higher end, once unobservable factors are controlled for. Importantly, by estimating a conditional FE logit model for each threshold value, any threshold-specific unobserved heterogeneity (labelled τ_{ik} in Section 3.1) is conditioned out; this allows us to present estimates that are unaffected by differences in reporting behaviour.

Estimated coefficients from the conditional FE logit models are presented in Columns (1) to (4) of Table 5. Consistent with previous results, vitality is the most important health dimension, with its coefficient roughly two-thirds larger than the next largest coefficient in each model. Also consistent with previous results, the effect of vitality is relatively homogeneous: its coefficient is only slightly larger at low SAH levels (4.166 for SAH>1) than at high SAH levels (3.525 for SAH>4). In contrast, the physical functioning health dimension has a much larger coefficient at low SAH levels (2.617 for SAH>1) than at high SAH levels (0.746 for SAH>4), suggesting that improvements in physical functioning affects individuals differently depending on whether they perceive their overall general health as poor or excellent. A similar result is found for social functioning and role physical. For the mental health dimension we find the reverse pattern, with a larger coefficient at high SAH levels (1.087 for SAH>4) than at low SAH levels (0.466 for SAH>1). In contrast to the physical functioning dimension, improvements in mental health appear to have a larger impact on SAH for individuals who are generally healthy than for individuals who are generally unhealthy.[13] This finding is consistent with the results seen in Columns (5) and (6) in Table 4, in which mental health is a more important determinant of SAH for the excellent health group than the poor health group.

Table 5: Logit and Linear Fixed-Effect Regression Models of Alternative Outcome Measures

	Conditional FE Logit Models				Linear FE Models	
	SAH > 1	SAH > 2	SAH > 3	SAH > 4	Health Satisfaction	Life Satisfaction
	(1)	(2)	(3)	(4)	(5)	(6)
Vitality	4.166**	3.866**	3.678**	3.525**	1.532**	0.654**
	(0.317)	(0.154)	(0.115)	(0.171)	(0.049)	(0.041)
Physical funct.	2.617**	2.403**	1.570**	0.746**	0.967**	0.067
	(0.230)	(0.120)	(0.099)	(0.147)	(0.048)	(0.035)
Bodily pain	1.731**	1.487**	1.588**	2.057**	0.776**	0.001
	(0.234)	(0.106)	(0.078)	(0.124)	(0.035)	(0.027)
Mental health	0.466	0.721**	1.023**	1.087**	0.525**	1.789**
	(0.312)	(0.166)	(0.132)	(0.205)	(0.058)	(0.054)
Role physical	0.951**	0.833**	0.489**	0.335**	0.447**	0.001
	(0.143)	(0.058)	(0.053)	(0.100)	(0.022)	(0.017)
Social funct.	1.517**	0.812**	0.205*	0.162	0.441**	0.324**
	(0.210)	(0.108)	(0.091)	(0.155)	(0.040)	(0.033)
Role emotional	-0.002	0.055	0.060	-0.147	0.029	0.130**
	(0.109)	(0.059)	(0.054)	(0.096)	(0.022)	(0.019)
Mean outcome	0.972	0.841	0.483	0.318	7.317	7.898
Individuals	1136	4368	7478	3986	16796	16796
Observations	8418	33223	57326	30178	104126	104104

Notes: Figures are estimated coefficients. Standard errors clustered at the individual-level are shown in parentheses. * and ** denote significance at .05 and .01 levels. All presented variables are scaled from 0 (worst) to 1 (best). Included in each model but not shown are the set of time-varying covariates shown in Table 1. SAH > j is a binary indicator representing a SAH score > j.

In Table 5 we also investigate whether our main conclusions are robust to the use of an alternatively worded general health question. In each wave of HILDA, respondents are asked to indicate on a scale of 0 to 10 how satisfied they are with their health. This question is analogous to overall life satisfaction questions and has been used, for example, to examine the impact of income on health (Frijters et al., 2005). The results in Column (5) show that the ordering and relative magnitudes of the health dimensions for the health satisfaction outcome closely mirror those found for the ordinal SAH measure.

[13] This finding also holds if we estimate conditional FE logit models with predicted mental health and physical health factors from a principal components analysis of the SF-36 health dimensions. The coefficient on the mental health factor is larger in the SAH>4 model, whereas the coefficient on the physical health factor is larger in the SAH>1 model.

In Column (6) of Table 5 we present equivalent estimates for the overall life satisfaction question ("All things considered, how satisfied are you with your life?"). Recently, a number of studies have estimated the effects of health on happiness and life satisfaction (e.g. Binder and Coad, 2013; Graham et al., 2011; Mukuria and Brazier, 2013; Powdthavee and van den Berg, 2011). Given our unique panel data, we are well-placed to add to this literature by examining whether the most important predictors of SAH are also the most important predictors of life satisfaction. Interestingly, we find that none of the physical health dimensions are statistically significant predictors. In contrast, the mental health dimensions – mental health (1.789), vitality (0.654), social functioning (0.324) and role emotional (0.130) – are strongly significant. If we omit the mental health dimensions from the model, the physical health dimensions become statistically significant with coefficients equalling roughly 0.3. This result suggests that physical health conditions are only important for life satisfaction if they affect mental health.

4. What is Self-Assessed Health Hiding?

The preceding analysis has shown that some health components are very important to an individual when they assess their health, while others are less consequential. Consequently, the impact of an economic treatment on SAH will depend upon the health components it affects. For example, it is more likely a researcher will find a statistically significant SAH effect if income affects vitality than if income affects bodily pain. This is perhaps an obvious point, but it is sometimes claimed that a certain variable has a zero health effect even when SAH is the sole measure used. In this Section we use an analysis of the income-health gradient to demonstrate how a reliance on SAH can mask important health effects.

4.1. Methodological Approach

The relationship between income and self-assessed health has been the subject of extensive research (for reviews see Gunasekara et al., 2011; Smith, 1999). This research is based on the supposition that income and other forms of socioeconomic status have large positive impacts on health – for example, through greater access to quality medical care. Surprisingly, however, most of the studies using either panel data models or instrumental variable (IV) models find weak income effects: some studies find quantitatively small but statistically significant effects (e.g. Contoyannis et al., 2004; Frijters et al., 2005; Jones and Wildman, 2008), while others find statistically insignificant effects (e.g. Apouey and Clark, 2010; Frijters and Ulker, 2008). A possible explanation is that income only positively affects health domains that are weakly associated with SAH. We examine this hypotheses by estimating instrumental-variable (IV) models of SAH, a binary indicator for having a long-term health condition, and the seven SF-36 health dimensions. We include the long-term health condition outcome because it is a commonly used health outcome measure alongside SAH.[14]

The use of an IV estimator is necessary given the strong likelihood of reverse causality and unobserved confounders. An IV estimator is also needed to overcome attenuation bias driven by classical measurement error in reported income (for a literature review see Moore et al., 2000). The instrumentation is particularly important in FE models because the attenuation bias arising from classical measurement error is known to be amplified if cross-sectional information is removed via fixed-effects. The first- and second-stage equations in our IV-FE model are given by:

$$\log(income_{it}) = \mu_{i1} + \lambda z_{it} + W_{it}'\pi_1 + u_{1it} \tag{5}$$

$$h_{it} = \mu_{i2} + \theta \log(income_{it}) + W_{it}'\pi_2 + u_{2it} \tag{6}$$

where $income_{it}$ is real annual household income from all sources, h_{it} is health (measured by SAH, presence of a health condition or one of the seven health dimensions), μ_i is an individual-level fixed-effect, and u_{it} is a random error term. W_{it} is a vector of characteristics that vary across individuals and time, including age (cubic function), educational attainment, marital status, children, and a number of life-event indicators that are potentially associated with income and health: a serious injury or illness to a family member or relative; the death of a family member, relative or friend; retirement from the workforce; being fired or made redundant; changing jobs; and receiving a promotion at work. The instrumental variable z_{it} is a binary variable formed from a question in waves 2-11 of HILDA. Individuals are asked whether they have experienced during the past year a "major financial improvement, e.g. won a lottery, received an inheritance". We observe 1,991 occurrences in our sample, with 14.3% of individuals reporting the event at least once.

Importantly, the IV reflects lottery wins and inheritances, but not other sources of windfall income: it is not statistically associated with the receipt of income from annuities;

[14] The health condition outcome variable is derived from the survey question "do you have any long-term health condition, impairment or disability (such as these) that restricts you in your everyday activities, and has lasted or is likely to last, for 6 months or more?" It is contained in each wave of the HILDA survey.

pension funds; workers compensation; accident or illness insurance; life insurance; redundancy or severance payouts; gifts from parents or other persons; or company shares, managed funds or property trusts (full results available upon request). This is important because other sources of windfall income, such as from accident or illness insurance, may be influenced by time-varying confounding variables, such as health shocks.[15]

Of course, lottery wins and inheritances are not randomly distributed across the population; their occurrence is often predicted by socioeconomic and demographic factors. We overcome this endogenous selection by including individual-level fixed-effects, which capture all time-invariant individual characteristics (such as cognitive ability, personality, time preference and risk aversion), and by including a range of time-varying covariates (such as education, marital status, and death of a relative). If this approach is sufficient we would not expect health this period to be a predictor of the IV in future periods. We test for such an association and find that SAH (t = 0.79), a long-term health condition (t = -0.08), the seventeen disaggregated long-term health conditions, vitality (t = 0.25), physical functioning (t = -0.71), bodily pain (t = 0.25), mental health (t = -1.79), social functioning (t = -1.01) and role emotional (t = 0.68) are all statistically insignificant predictors of the IV. Of the 26 estimated models, the only significant predictor was role physical (t = 2.14). Overall, these results suggest that the IV exogeneity assumption is valid in our context.

4.2. Results

Table 6 presents OLS and IV-FE models of SAH, presence of a health condition, and the seven SF-36 health dimensions, which are all scaled to range between 0 (worst health state) and 1 (best health state) to ease interpretation. For brevity, only the coefficient estimates for log annual household income are shown. Estimates for the first-stage equation are also omitted from Table 6, with only the F-statistic on the IV presented. Importantly, the effect of the IV on log income in each model is highly significant, and suggests that an inheritance or lottery win increases log income by 0.21 units for both males and females.

As is typically found, log income is a highly significant predictor of health in OLS models. For example, column (1) suggests that a one-unit change in log income increases SAH (scaled to range between 0 and 1) by 0.055 units (t-statistic = 13.8), and decreases the probability of a long-term health condition by 9.5 percentage-points (t-statistic = -13.3). It is

[15] Given the illustrative nature of our analysis, an attractive feature of using lottery wins and inheritances as an IV is that it is one of the most commonly used sources of exogenous income variation in health economics – examples include Apouey and Clark (2010), Gardner and Oswald (2007), Lindahl (2005), Kim and Ruhm (2012), Meer et al. (2003), and Michaud and Van Soest (2008).

usually argued that OLS estimates are positively biased (negatively biased in the case of the health condition outcome), because of reverse causality and unobserved confounders. Though, the OLS estimates may be negatively biased due to measurement error in income. It is also possible that the magnitude and sign of the endogeneity bias will differ across the health dimensions. For instance, the OLS models of physical functioning may suffer from greater reverse causality than the OLS models of vitality, if impaired physical functioning reduces labour supply and wages to a greater extent than vitality does. Similarly, the extent of endogeneity bias may differ across genders if reverse causality is more important for men because of their higher labour force attachment and greater likelihood of working in physically demanding occupations.

Table 6: Estimated Effects of Log Income on Health from OLS and IV-FE Models

	Males		Females	
	OLS	IV-FE	OLS	IV-FE
Dependent variable	(1)	(2)	(3)	(4)
SAH	0.055**	0.021	0.049**	0.018
	(0.004)	(0.021)	(0.004)	(0.020)
Health condition	-0.095**	-0.017	-0.085**	-0.078
	(0.007)	(0.048)	(0.007)	(0.043)
Vitality	0.038**	0.008	0.029**	0.028
	(0.003)	(0.018)	(0.003)	(0.018)
Physical functioning	0.049**	0.014	0.045**	0.005
	(0.004)	(0.017)	(0.004)	(0.016)
Bodily pain	0.054**	-0.019	0.044**	0.014
	(0.004)	(0.022)	(0.004)	(0.022)
Mental health	0.032**	0.015	0.029**	0.052**
	(0.003)	(0.016)	(0.003)	(0.016)
Role physical	0.089**	-0.048	0.072**	0.115**
	(0.006)	(0.036)	(0.005)	(0.038)
Social functioning	0.059**	-0.008	0.052**	0.063**
	(0.004)	(0.024)	(0.004)	(0.024)
Role emotional	0.066**	-0.055	0.056**	0.085*
	(0.005)	(0.036)	(0.005)	(0.037)
1st-stage F-statistic	-	120.70	-	161.11
Individuals	5190	5190	5640	5640
Observations	28601	28601	31728	31728

Notes: Each cell represents estimated coefficients from a separate regression. Standard errors clustered at the individual-level are shown in parentheses. All dependent variables range from 0 to 1. * and ** denote significance at .05 and .01 levels. Included in each model but not shown are the set of time-varying covariates shown in Table 1 and indicators of a serious injury or illness to a family member or relative; the death of a family member, relative or friend; retirement from the workforce; being fired or made redundant; change jobs; and receiving a promotion at work.

The IV-FE estimates in Columns (2) and (4) indicate that income is statistically insignificant in the SAH and health condition models, for both males and females. As discussed in Section 4.1, this finding is not uncommon. For these general health outcomes the IV-FE estimates are smaller than the OLS estimates, suggesting that positive endogeneity bias exists. The male IV-FE estimates are also statistically insignificant for each of the SF-36 health dimensions. Given the broad coverage of these measures, the set of male results indicates that income does not causally impact the health of men.

In contrast to the male results, the female estimates for mental health (0.052), role physical (0.115) and social functioning (0.063) are statistically significant at the 1% level, and the estimate for role emotional (0.084) is statistically significant at the 5% level. These estimates suggest that an increase in female income increases feelings of peacefulness and happiness (mental health), reduces limitations with work or other activities due to physical ill-health (role physical), reduces the interference of ill-health with social activities (social functioning), and reduces limitations with work or other activities due to emotional ill-health (role emotional). Ex post these findings are not surprising. It is quite plausible that increased income improves unhealthy individuals' ability to carry-out day-to-day activities and to interact socially (e.g. by expanding the range of feasible transport options for those less physically mobile). In contrast, it is harder to conceive of ways in which income could substantially improve an individual's level of vitality.

Most importantly, given the context of this analysis, these large and significant effects for females are not reflected in the SAH or health condition coefficients, and would typically be 'missed' given that detailed health information is relatively rare in nationally representative longitudinal surveys. From our findings in Section 3, a likely explanation for the absent SAH effect is that mental health, role physical, social functioning and role emotional are only weakly associated with SAH, and therefore changes in these dimensions will only weakly cause changes in SAH. Note that the absent SAH effect for females is not the result of using the ordinal SAH outcome in a linear model. If we instead model binary indicators of poor health and excellent health, as is often done, we also find statistically insignificant IV-FE estimates. Similarly, the IV coefficient is statistically insignificant in a FE ordered logit model of SAH (i.e. reduced form equation).

The observed health effects are also missed if we additionally use disaggregated health conditions, which is a common practice. The HILDA data contains 17 long-term health condition categories, such as "hearing problems", "limited use of feet or legs", "chronic or recurring pain", and "a nervous or emotional condition which requires treatment".[16] For all 17 outcomes, the IV-FE income estimate is small and statistically insignificant (full results available upon request). These estimates reinforce the conclusion that important income effects are not captured with commonly collected health information.

5. Conclusion

Many studies rely on self-assessed general health (SAH) as a global measure of health status. This has two important consequences. First, it is difficult to interpret the estimated health effects of economic interventions, because SAH provides no guidance to researchers regarding the components of health that are being affected. For example, if socioeconomic status is found to improve SAH, is this because SES reduces pain levels, improves physical functioning, or is some other process at work? Clearly, the answer to this question provides important insights. The second consequence of relying on SAH is that important health effects can be missed, as many aspects of health are poorly captured by SAH.

In this paper we investigate the first issue and demonstrate the second. We use longitudinal data that includes a detailed health instrument, and estimate linear, ordered-logit, dynamic and instrumental-variable fixed-effects models. Our first main conclusion is that when an individual assesses their health, the most important consideration is whether they are feeling full of life and energetic (vitality). This result is robust to the use of alternative estimators, sub-groups and general health outcome measures. The next most important considerations are whether ill health is limiting mobility (physical functioning) and causing pain (bodily pain). More broadly, mental health and physical health are roughly equal considerations; though, physical health appears to be more important for individuals who are generally unhealthy, while the reverse is true for mental health. Our second main conclusion is that large and significant income-health effects can be missed if researchers rely on SAH and other general health measures. We find for females that income significantly improves mental health, and reduces problems in carrying out day-to-day activities due to ill-health, but not SAH or the likelihood of having a long-term health condition.

To the best of our knowledge this is the first economics paper that rigorously examines the self-assessed health variable, which is surprising given the attention paid to

[16] The 17 health conditions categories are: sight problems; hearing problems; speech problems; blackouts, fits or loss of consciousness; difficulty learning things; limited use of arms or fingers; difficult gripping things; limited use of feet or legs; nervous or emotional condition which requires treatment; any condition that restricts physical activity; any disfigurement or deformity; any mental illness which requires help; shortness of breath; chronic or recurring pain; effects of a head injury, stroke or brain damage; condition which is restrictive even though it is treated; any other condition.

other commonly used metrics. For example, there are several papers that have examined the constituent parts of job satisfaction (Clark, 2001) and life satisfaction (van Praag et al., 2003). A natural extension to our study would be to investigate whether the same health components (i.e. vitality, physical functioning and pain) are driving SAH responses in other countries. Another extension would be to determine whether there are important predictors of SAH not contained in our data, and therefore not highlighted in this paper. However, these extensions would require nationally representative panel data that includes repeated measures of SAH and detailed disaggregated measures of health. We are unaware of such datasets.

An important implication of our findings is that particular care must be taken when interpreting econometric models of SAH, because a zero SAH effect does not imply zero effects in all components of health. As stated in Apouey and Clark, 2010, "health is not a holistic concept, and we need to both be clear about what kind of health we are talking about, and be ready for the possibility that different types of health behave in very different ways" (p. 22). To gain a clearer understanding of population health effects, it is important for large, nationally representative surveys, which often only contain SAH, to more frequently include detailed health questionnaires.

References

Almond D, Mazumder B. The 1918 influenza pandemic and subsequent health outcomes: an analysis of SIPP data. American Economic Review 2005; 258-262.

Andersen M, Lobel M. Predictors of Health Self-Appraisal: What's Involved in Feeling Healthy? Basic and Applied Social Psychology 1995;16; 121-136.

Apouey B, Clark A. Winning big but feeling no better? The effect of lottery prizes on physical and mental health. IZA Discussion Paper No. 4730. 2010.

Arellano M, Bond S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. The Review of Economic Studies 1991;58; 277-297.

Bago d'Uva T, O'Donnell O, van Doorslaer E. Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. International Journal of Epidemiology 2008;37; 1375-1383.

Baron-Epel O, Kaplan G, Haviv-Messika A, Tarabeia J, Green MS, Nitzan Kaluski D. Self-reported health as a cultural health determinant in Arab and Jewish Israelis:: MABAT--National Health and Nutrition Survey 1999-2001. Social science & medicine 2005;61; 1256-1266.

Bateman IJ, Mawby J. First impressions count: interviewer appearance and information effects in stated preference studies. Ecological Economics 2004;49; 47-55.

Benyamini Y, Idler EL, Leventhal H, Leventhal EA. Positive Affect and Function as Influences on Self-Assessments of Health. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 2000;55; P107-P116.

Binder M, Coad A. “I'm afraid I have bad news for you…” Estimating the impact of different health impairments on subjective well-being. Social science & medicine 2013.

Blundell R, Bond S. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 1998;87; 115-143.

Bound J. Self-Reported versus Objective Measures of Health in Retirement Models. Journal of Human Resources 1991;26; 106-138.

Case A, Paxson C. Height, health, and cognitive function at older ages. The American Economic Review 2008;98; 463-467.

Clark AE. What really matters in a job? Hedonic measurement using quit data. Labour Economics 2001;8; 223-242.

Contoyannis P, Jones AM, Rice N. The dynamics of health in the British Household Panel Survey. Journal of Applied Econometrics 2004;19; 473-503.

Etilé F, Milcent C. Income-related reporting heterogeneity in self-assessed health: evidence from France. Health Economics 2006;15; 965-981.

Ettner SL. New evidence on the relationship between income and health. Journal of health economics 1996;15; 67-85.

Foster G, Frijters P, Johnston DW. The triumph of hope over disappointment: A note on the utility value of good health expectations. Journal of Economic Psychology 2012;33; 206-214.

Frijters P, Haisken-DeNew JP, Shields MA. The causal effect of income on health: Evidence from German reunification. Journal of health economics 2005;24; 997-1017.

Frijters P, Ulker A. Robustness in health research: Do differences in health measures, techniques, and time frame matter? Journal of health economics 2008;27; 1626-1644.

Graham C, Higuera L, Lora E. Which health conditions cause the most unhappiness? Health Economics 2011;20; 1431-1447.

Groot W. Adaptation and scale of reference bias in self-assessments of quality of life. Journal of health economics 2000;19; 403-420.

Gunasekara FI, Carter K, Blakely T. Change in income and change in self-rated health: Systematic review of studies using repeated measures to control for confounding bias. Social science & medicine 2011;72; 193-201.

Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. Journal of Health and Social Behavior 1997; 21-37.

Johnson RC. The Health Returns of Education Policies from Preschool to High School and Beyond. The American Economic Review 2010;100; 188-194.

Jones AM, Rice N, Roberts J. Sick of work or too sick to work? Evidence on self-reported health shocks and early retirement from the BHPS. Economic Modelling 2010;27; 866-880.

Jones AM, Schurer S. How does heterogeneity shape the socioeconomic gradient in health satisfaction? Journal of Applied Econometrics 2011;26; 549-579.

Jones AM, Wildman J. Health, income and relative deprivation: evidence from the BHPS. Journal of health economics 2008;27; 308-324.

Jürges H. True health vs response styles: exploring cross-country differences in self-reported health. Health Economics 2007;16; 163-178.

Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. Social science & medicine 2009;69; 307-316.

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

Krause NM, Jay GM. What do global self-rated health items measure? Medical Care 1994; 930-942.

Lindeboom M, Van Doorslaer E. Cut-point shift and index shift in self-reported health. Journal of health economics 2004;23; 1083-1099.
Manderbacka K. Examining what self-rated health question is understood to mean by respondents. Scandinavian Journal of Public Health 1998;26; 145-153.
McGarry K. Health and Retirement Do Changes in Health Affect Retirement Expectations? Journal of Human Resources 2004;39; 624-648.
Meer J, Miller DL, Rosen HS. Exploring the health-wealth nexus. Journal of health economics 2003;22; 713-730.
Moore JC, Stinson LL, Welniak EJ. Income measurement error in surveys: A review. Journal of Official Statistics 2000;16; 331-362.
Mukuria C, Brazier J. Valuing the EQ-5D and the SF-6D health states using subjective well-being: A secondary analysis of patient data. Social science & medicine 2013;77; 97-105.
Powdthavee N, van den Berg B. Putting different price tags on the same health condition: Re-evaluating the well-being valuation approach. Journal of health economics 2011;30; 1032-1043.
Ryan RM, Frederick C. On Energy, Personality, and Health: Subjective Vitality as a Dynamic Reflection of Well-Being. Journal of Personality 1997;65; 529-565.
Shmueli A. Socio-economic and demographic variation in health and in its measures: the issue of reporting heterogeneity. Social science & medicine 2003;57; 125-134.
Silles MA. The causal effect of education on health: Evidence from the United Kingdom. Economics of Education Review 2009;28; 122-128.
Singh-Manoux A, Martikainen P, Ferrie J, Zins M, Marmot M, Goldberg M. What does self rated health measure? Results from the British Whitehall II and French Gazel cohort studies. Journal of Epidemiology and Community Health 2006;60; 364-372.
Smith JP. Healthy bodies and thick wallets: the dual relation between health and economic status. The Journal of Economic Perspectives 1999;13; 145-166.
Tarlov AR, Ware JE, Jr., Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study: An Application of Methods for Monitoring the Results of Medical Care. JAMA 1989;262; 925.
Taylor MP. Tell me why I don't like Mondays: investigating day of the week effects on job satisfaction and psychological well‐being. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2006;169; 127-142.
van Praag BMS, Frijters P, Ferrer-i-Carbonell A. The anatomy of subjective well-being. Journal of Economic Behavior & Organization 2003;51; 29-49.
Ware JEJ. SF-36 health survey update. Spine 2000;25; 3130-3139.