Does self-assessed health measure health?[1]

Denise Doiron[a], Denzil G. Fiebig[a], Meliyanni Johar[b*] & Agne Suziedelyte[c]

[a] School of Economics, University of New South Wales, Australia

[b] Economics Discipline Group, University of Technology Sydney, Australia

[c] Centre for Health Economics, Monash University, Australia

Abstract

Despite concerns about subjective measures of health, the ubiquitous self-assessed health (SAH) remains the measure of health most used by researchers, in part reflecting its ease of collection and in part the observed correlation between SAH and objective measures of health. Using a unique Australian data set, which consists of survey data linked to administrative individual medical records, we present empirical evidence demonstrating that SAH indeed predicts future health, as measured by hospitalizations, out-of-hospital medical services and prescription drugs. The effects of SAH decrease substantially once we control for past utilization which indicates that a large proportion of the variation in SAH reflects variation in actual health. Nevertheless, SAH has extra predictive power implying that health perceptions of individuals have additional health content that influences future health. Our large sample size allows very

---

[1] This research is funded by an ARC Discovery Project grant (DP110100729). It uses data from the 45 and Up Study, which is managed by the Sax Institute in collaboration with major partner Cancer Council New South Wales (NSW), and partners the Heart Foundation (NSW Division); NSW Ministry of Health; *beyondblue*; Ageing, Disability and Home Care, NSW Family and Community Services; Australian Red Cross Blood Service; and UnitingCare Ageing. This project was undertaken by the University of Technology Sydney and utilized Pharmaceutical Benefit Schedule (PBS) data supplied by the Department of Human Services and linked to the 45 and Up Study by the Sax Institute. Data linkage for the hospital admission data used in the project was undertaken by the Centre for Health Record Linkage. The 45 and Up Study has the approval of the University of NSW Health Research Ethics Committee; this project has ethics approval from the NSW Population and Health Services Research Ethics Committee and the Department of Health and Ageing Departmental Ethics Committee. The study's findings are those of the authors' and do not necessarily represent the views of the Department of Health and Ageing, or the Department of Human Services.

* Corresponding author: PO Box 123, Broadway NSW 2007, Australia. Tel.+61 2 9514 7722. Fax +61 2 9514 7722. Email address: meliyanni.johar@uts.edu.au

disaggregate analysis and we find that SAH predicts more serious, chronic illnesses better than less serious illnesses.

Running title: Does self-assessed health measure health?

Keywords: health; self-assessed health; health care use; administrative data.

JEL codes: I1

"*I'm not sure what self-assessed health measures but I do know it's not health*"

anonymous

1. Introduction

Asking people to rate their health is a pervasive question in health surveys. Such a question is easily asked and the resulting self-assessed health (SAH) variable is routinely used by researchers in analyses across many disciplines. In fact Jylhä (2010) contends:

"Very few indicators attract the interest of researchers from so many different fields and disciplines".

While there is scepticism about the use of subjective health measures rather than more objective measures (Bound, 1991 and Crossley and Kennedy, 2002), the former are much more readily available to researchers. In addition, there exist robust findings of positive correlations between SAH and actual health and mortality; see for example Idler and Benyamini (1997), McCallum et al (1994) and Van Doorslaer and Gertham (2003). Consequently the use of SAH as a measure of health remains ubiquitous.

Despite its popularity, concerns remain about the validity of the empirical regularities involving SAH, so it is natural and appropriate to ask what SAH actually measures. See for example Layes et al (2012) and references therein. Our overall aim is to contribute to this literature and in particular to better understand to what extent SAH represents actual health. The notion of "true" health is itself an elusive and complex concept that is likely to be multidimensional. As such, our empirical approach involves an investigation of whether SAH predicts a range of future health outcomes. Furthermore, the richness of our data allows us to disaggregate these health outcomes into specific illnesses thereby testing if SAH predicts certain illnesses better than others. The data are derived from a large sample survey linked with multiple years of comprehensive health administrative data sets, covering hospital admissions, out-of-hospital medical services and prescription drugs.

The empirical strategy involves the use of prospective models of future health care utilizations as measures of health outcomes. Current SAH is included as a covariate, and the significance of its coefficients would indicate that SAH has “health content”, meaning that it is predictive of future health. As we also seek to understand the source of the health content of SAH, we start with a baseline model that includes SAH and an extensive set of control variables. We then add a subjective quality of life (QoL) variable. This is to satisfy those sceptics who believe that subjective survey variables like SAH merely capture the respondents’ reporting style. We then further augment the model with past utilizations, interpreted as negative shocks to health. If SAH captures actual health, we expect the estimated coefficients on the SAH variables to become smaller and potentially disappear.

We find that SAH substantially increases the utilization of health care services, and this is robust to the inclusion of QoL variables. The effects of poor SAH vary from 23% to 28% relative to the mean across different health care services. When we control for past utilization, the SAH impact is reduced, especially for recent shocks. We take all of these results as convincing evidence that SAH does have actual health content.

2. Literature review

The literature around SAH can be categorized into two streams: one that seeks to explain the variation in SAH and another that tests the usefulness of SAH. The focus of the first stream has been on the potential bias in responses to the subjective SAH question with candidate explanations including discrepancies due to cultural norms, individual social status and health history, and different appraisals of a given clinical health condition.

Two common approaches that have been used to explore the reporting bias have been to study the variation in SAH beyond that explained by more objective health indicators and a vignette approach to fix or anchor a health state (that is, freeing it from individual subjectivity and expectation). Applying the first approach, lower income individuals are found more likely to report a poor level of SAH than higher income groups (e.g., Humphries and van Doorslaer, 2000; Hernandez-Quevedo et al., 2004; Etile and Milcent, 2006), women are more optimistic than men and are less likely to report poor health (Arber and Cooper, 2006), and people of certain nationalities are more optimistic in rating their health than others (Jürges, 2007). Using vignette anchoring, Salomon et al. (2004) also show that self-reported mobility suffers from cross-country bias, Bago d'Uva et al. (2008) find that highly educated individuals are more pessimistic about health and rate a health state ‘poor’ more often, while Lindeboom and van Doorslaer (2004) find significant reporting heterogeneity associated with age and sex. Recently Layes et al. (2012) use another preference-standardized measure to elicit individual true health valuation. Their results are consistent with the vignette-based studies, particularly SAH responses exhibit age-norming: SAH overestimates the health status of the elderly because optimism in reporting behaviour increases with age.

Related to this stream of research is the literature on reporting error in health-related survey questions. Crossley and Kennedy (2002) use a natural experiment where some respondents answer the SAH question twice to show that SAH can be unreliable. They show that SAH responses vary depending on whether the question is asked before or after a series of objective health questions suggesting a learning effect. Other sources of errors are related to a lack of awareness of health problems (e.g., Johnston et al., 2009) and the presence of incentives for misreporting; for example the misreporting of disability status in order to qualify for disability payments or the concealment of chronic illness to avoid labour market disadvantage (e.g., Gupta and Jurges, 2012 and references within).

This study is more closely related to the second stream of research: can SAH predict future illnesses? Which ones? Several studies have focused on the predictive power of SAH on survival (early papers include Idler and Benyamini (1997) and McCallum et al. (1994)). Using a sample of Swedish adults, van Doorslaer and Gerdtham (2003) find that SAH predicts mortality, especially for younger individuals. Huisman et al. (2007) find that SAH's predictive power for mortality varies by sex and education. Specifically, SAH predicts mortality better for men with tertiary education than men with low education. Jylhä (2009) discusses pathways in which SAH may predict mortality. Lee (2000) finds that SAH predicts not only mortality, but also functional decline (physical and instrumental) among Americans aged 70 and over. In a 16 year follow-up study involving more than a thousand Danish men, Møller et al. (1996) find that, even after controlling for risk factors and other potential confounders, those reporting poor and miserable SAH have 6.5 and 18.6 fold higher risk of fatal and non-fatal coronary heart disease, respectively, than those reporting extremely good SAH. Using British data, Manor et al. (2001) find that SAH predicts more serious conditions like heart diseases, cancer and diabetes better than less serious conditions like high blood pressure, migraine, eczema and hay fever.

Relatively few studies examine the impact of SAH on health care utilization (Connelly et al. 1989; Miilunpalo et al., 1997; Long and Marshall 1999). Using the same survey data as the one used in this study, Johar et al. (2013) find a strong negative SAH gradient in presentations to an emergency department: those reporting poor health are more likely than others reporting better health to present to an emergency department, and they also tend to make more presentations. Finally, Ellis et al. (2013) show that SAH has strong predictive power over aggregate health care expenditure.

We contribute to this literature in two ways. First, our objective health measures are derived from administrative individual medical records, which cover a broad range of medical services including GP and specialist visits, diagnostic services, prescription drugs, and hospitalizations. These medical services are further disaggregated by specific diseases. The use of administrative data minimizes measurement error, particularly due to recall/reporting bias.

Additionally, under Australian universal free health care system, the bias arising from failing to capture sick people who do not seek medical treatment is likely to be minimal. Most of the above mentioned studies use self-reported illnesses and physical health scores, such as SF-36, as measures of objective health, which may be reported with errors. Additionally, the large size of our sample makes it possible to identify and credibly examine relationships between SAH and relatively rare illnesses, such as cancer. Most of the other studies in this area have been based on relatively small sample sizes. Our second contribution is the use of a prospective model, which avoids the issue of reverse causality. In a contemporaneous model, which is usually estimated in the literature, it is difficult to disentangle the effect of SAH on objective health from the effect of health shocks on SAH.

3. Method

The aim of this study is to test the health content of SAH. Our identification strategy relies on the timing of events (prospective models) and the inclusion of an extensive set of controls to minimize the impact of any remaining reverse causality and the presence of confounding factors. We firstly estimate multivariate models of prospective health care utilization as a function of SAH, with extensive control variables. These models will confirm whether SAH has independent effects on prospective health outcomes. The use of extensive controls improves the comparability of SAH across individuals by reducing reporting errors and biases due to differences in health-threshold levels and reporting norms across subgroups. Studies have shown that these biases can mostly be explained by individual background characteristics, such as demographic characteristics, origins and socioeconomic status (e.g. Lindeboom and Van Doorslaer, 2004).

We then sequentially add self-reported quality of life (QoL) and past utilization. Like SAH, QoL is a subjective survey variable influenced by reporting norms and biases. Including both SAH and QoL variables in the model can shed light on whether SAH merely reflects the reporting style of respondents. If responses to any subjective variable in a survey simply reflect variation in the respondents' personality or psychological type, such as the respondent being a naturally anxious or conscientious person, the size of the SAH coefficients may diminish. We expect SAH coefficients to be further reduced when past utilization is added. In the rest of the regressions, we include QoL as a control for variations in personality types, but past utilization is excluded as we want SAH to capture all actual health aspects.

As an additional check, we restrict attention to a homogenous sub-sample of relatively healthy individuals and we compare the predictive power of SAH for the sub-group of these healthy individuals who experience a negative health shock in the form of a hospital admission in the base period with the remaining healthy respondents. For this experiment, an individual is defined as relatively healthy if he/she was not admitted to hospital and not diagnosed with a chronic condition up until two years before the base period. If SAH is measuring actual health, then in any category of SAH there is likely to be a group with stable health and a group who are experiencing worsening health and hence are more likely to have transitioned into their current SAH level from a better level. Thus, we expect the effect of SAH on prospective health outcomes to be stronger for individuals who experience a health shock, as it is picking up both the level of health and its change.

Next, we compare the predictive power of SAH across illness groups. The illness groups are defined by collating medical codes and service numbers of inpatient and outpatient utilizations. Separate models which control for QoL are estimated for each illness group. Of particular interest are results focussing on diseases related to mental disorders since the literature to date has not examined such cases. Finally, we explore heterogeneity in the remaining health content of SAH along the dimensions of age, sex and education by including their interaction terms with SAH.

4. Data

The data are derived from four data sets. The first source is the 45 and Up Study, which is a cross-section survey of non-institutionalized individuals aged 45 and over (45+) in the state of New South Wales (NSW), Australia. NSW is the most populous state in Australia with a population of about 7.3 million, 39% aged over 45. The 45 and Up Study consists of over 267,000 respondents, surveyed once during 2006–2010, with the largest collection taking place in 2008 (about 80%). The variation in survey years is part of the data collection design and is not a choice for respondents. A random selection of persons within the 45+ population is chosen from the Medicare Australia database for the survey. People over 80 years of age or resident in rural and remote areas were oversampled. The Medicare database covers everyone who has access to public health insurance (basically all permanent residents in Australia). The survey collects extensive information about the respondents' current health status, quality of life and history of own, parents' and siblings' chronic illnesses, as well as demographic and socio-economic characteristics.

The 45 and Up Study can be linked to multiple health administrative data sets at the respondent level. Only a few survey-administrative data linkages of this scale exist anywhere in the world and it is unique in Australia. The three administrative databases used in this study are: the NSW Admitted Patient Data Collection (APDC), the Medicare Benefits Schedule (MBS) and the Pharmaceutical Benefits System (PBS). The APDC data includes all hospital separations by the survey respondents during 2000-2009; the MBS data consist of out-of-hospital medical services for which a Medicare subsidy was paid and the PBS data of prescription drugs for which a Medicare subsidy was paid during 2006-2009. About 80% of prescription drugs dispensed in Australia are subsidized.

To apply a one-year prospective framework and to have available past health care utilization, we focus on the survey respondents who completed the survey in 2007 and 2008 (226,121 observations). A few respondents with invalid age and sex or who were unsolicited for the survey were excluded. To deal with missing data, we computed the percentage of observations with missing values for each variable in the models. If more than 1% of observations had missing values, the observations were kept and dummy variables for missing information were added to the model. If less than 1% of observations had missing values, the observations with missing values were deleted. Our final analysis sample consists of 212,574 observations (80% of the original sample). We note that in order to conduct the analysis at the level of disaggregation we use, large samples are crucial.

In the health care utilization models, separate logit regressions are estimated for 5 dependent variables. These are binary variables defined over the 12 months following the survey date; the calendar time covered by these variables is specific to the individual in that it depends on the date at which the respondent completed the survey. The utilization variables are:

1. *Hospital*, a binary variable equal to 1 if the respondent had at least one hospital admission (private or public);
2. *GP*, a binary variable equal to 1 if the respondent had more than 6 normal-hour GP consultations;
3. *Specialist*, a binary variable equal to 1 if the respondent had at least one specialist visit;
4. *Other medical*, a binary variable equal to 1 if the respondent had more than 10 of any of the following non-specialist out-of-hospital medical services: haematology, psychology, 0, pathology, physiotherapy, podiatry and radiation oncology;
5. *Drugs*, a binary variable equal to 1 if the respondent consumed more than 2 types of drugs.

For all 5 variables, the threshold used to switch the dummy variable to 1 corresponds to the sample median.

In the case of prescription drugs we use the first digit of the Anatomical Therapeutic Codes (ATC) to define drug groups or types of drugs. We focus on drug groups for several reasons. A consumption of multiple drug groups indicates comorbidities. Also, the use of drug groups rather than individual drugs minimises potential bias due to coverage of PBS data on drugs for which a subsidy was paid only. It is very unlikely that all drugs within a drug group are not subsidized.

We gather the information from hospital diagnoses (International Classification of Disease version 10 codes of primary diagnosis), specialities of specialist visits (MBS item numbers) and ATC drug groups to define 14 major illness groups for the illness models. Table 1 details this mapping, illustrating the scope of our analysis. As for the utilization outcomes, prospective models are used; specifically, illnesses are defined by binary variables measuring the incidence of treatment for the illness in the 12 months following the survey date. Note that as for utilization variables, the 14 outcomes representing illnesses are not mutually exclusive as an individual with comorbidities (across the illness groups) will have a dependent variable equal to 1 for more than one illness outcome. Separate logit regressions are estimated for the 14 illness groups.

[Insert Table 1]

The SAH information is obtained from the survey data. This variable is based on a five-point scale answer to the question “In general, how would you rate your health?” The scale reflects the 5 possible choices: excellent, very good, good, fair and poor. Since there are only a small number of respondents who rate their health as poor (less than 5%), we combine fair and poor responses together representing the unhealthiest group of sample respondents. At the other end of the spectrum, the healthiest group consists of those reporting excellent or very good SAH. The sample proportions in the three SAH groups are: 50.7%

excellent or very good, 32.6% good and 13.5% fair or poor (3.2% of the sample did not report their SAH). Similarly, respondents were asked to rate their QoL "How would you rate your quality of life?" on the 5-point scale from excellent to poor.

The survey data contain extensive information about the respondents' demographic (age, sex, residential location, marital status, country of birth, language etc.) and socio-economic characteristics (education, income, employment, health insurance, housing), lifestyle (smoking, alcohol consumption, body weight), daily health limitations (physical functioning and mental distress), as well as family history of illness (parents' and siblings' illnesses). In total there are 103 controls that are included in our analysis to capture the various categories of individual backgrounds. Some of the control variables are likely to correlate with aspects of health (e.g. body weight, physical distress) and their inclusion is likely to reduce the effects of SAH on future utilization and illness; in this sense, our results can be treated as lower bounds for the true predictive power of SAH.

For past utilization, we use the historical aspect of the linked administrative data. In effect, we are including a vector of lagged dependent variables although the time frame covered by these variables is in some cases longer than the one-year-ahead time frame used for the dependent variables. Specifically, we add hospital diagnoses for all hospital admissions in the past five years and out-patient service and prescription drug use in the last 12 months from the survey date (128 variables). We additionally estimate a specification, in which the utilization of all health care services in the past year is disaggregated into utilization in the past four quarters, as it is expected that the correlation between the more recent illnesses and SAH would be stronger.

5. Results

5.1.Descriptive statistics

Table 2 presents the means and sample proportions of the main control variables by the self-assessed health (SAH) status. As expected there is positive correlation between SAH and quality of life. SAH decreases with age. People in worse SAH are more likely to suffer from chronic conditions, psychological distress, physical activity limitations and disability and engage in unhealthy lifestyle (except for alcohol consumption). There is also variation in SAH by socio-economic status, as measured by education, household income, and an index of relative socioeconomic advantage measured at the local area level (SEIFA), with more socio-economically advantaged individuals tending to report better health.

[Insert Table 2]

Means of the outcome variables by SAH status are presented in Table 3. Worse SAH is positively associated with the utilization of all types of health care services, especially drugs, in the 12 months following the survey date. The incidence of all the illnesses in the next 12 months is also higher among the respondents in worse health.

[Insert Table 3]

5.2.Health content of SAH

Table 4 reports our main results. Panel A shows that there is a significant positive relationship between worse SAH and future health care utilization even after controlling for extensive demographic and socio-economic characteristics. The presented figures are average partial effects from the logit regressions. Worse SAH is found to increase utilization of all major health care services in the next 12 months, although there is some variation in the effects across services. The finding that SAH predicts a variety of health care services and not only visits to GPs who are gatekeepers in the Australian health care system, does not support the common scepticism that SAH merely captures "worried well" individuals who need reassurances from their doctor. Relative to those with very good health, those in slightly worse health have 3 percentage points higher probability of hospitalization and 6-7 percentage points higher probability of out-of-hospital services utilization, while those in fair/poor health have 7 percentage points higher probability of hospitalization and 10-11 percentage points higher probability of out-of-hospital services utilization. The gradient in the SAH effects is clear and is a pervasive feature of the results to follow.

In order to place the size of the effects of poor SAH in perspective, we note that these impacts are as large as or larger than the effects of pre-existing medical conditions such as heart disease, stroke and diabetes, which are considered to be more objective health indicators. Relative to the mean utilization rates, the effects of poor SAH are largest on hospitalizations (27.7% relative to the mean of 0.256) and prescription drug use (26.9% relative to the mean of 0.406).

As a sensitivity check on the decision to discretise the health outcomes into binary variables, we employ an alternative modelling strategy where latent class negative binomial models are estimated for the number of GP visits, drugs, and other medical services consumed. The results of these models are consistent with those presented in Table 4. For all dependent variables and all two-class latent models, we find that worse SAH increases utilization of health care services in the next period, and that these effects are statistically and economically significant. For example, people in poor SAH are found to have 26.3% (class 1, 'low utilization') to 50.7% (class 2, 'high utilization') more drug prescriptions than people in excellent or very good SAH.

[Insert Table 4]

Adding the self-rated QoL measure (panel B of Table 4) hardly changes the results in either precision or size of coefficients, confirming that SAH does have health content, and is not simply capturing a reporting style or other psychological factor. In fact, the average partial effects of SAH slightly increase (by 1%-7%). It also shows that QoL does not have additional health content once SAH is included in the model. Poor QoL itself slightly reduces future utilization of all health care services, with the exception of hospitalizations.

Results presented in Table 5 support the contention that SAH represents actual health, in the sense that variations in SAH capture variations in past health care services utilization. For convenience, Panel A replicates the results of the baseline model. Panel B reports the results based on models that include past utilization. They show that the effects of worse SAH on future health care utilization remain significant when past utilizations representing negative health shocks are added to the model, but their sizes are dramatically reduced. This pattern suggests that a substantial proportion of the variation in SAH is explained by the variation in actual health, although there is still health content in the perceptions about general health of individuals since SAH still has significant independent predictive power for future health. In panel C, the utilization of all services in the past year is disaggregated into the utilization in the past 4 quarters (hospitalizations 5, 4, 3, and 2 years ago are still included, but not disaggregated into quarters). The effects of SAH in this model decrease further, indicating that SAH is affected differentially by the timing of recent health events, especially in the case of out-patient services and prescription drug use. Compared to the baseline model estimates (Panel A), the estimated effects of SAH decline by 58%-88% once the past utilization measures are included (Panel C).

[Insert Table 5]

Table 6 presents results for the subset of respondents classified as "healthy". Close to 18% of the sample are defined as being relatively healthy according to our definitions (described in Section 3). As expected, the proportion of individuals in excellent or very good health is higher in this subsample (65.36%) than in the rest of the sample (48.37%). Around 12% of the "healthy" individuals had a health shock (a hospitalization) in the base period. The proportion of individuals in excellent or very good SAH is lower among the hospitalized respondents (60.43%) than among those who did not have a hospitalization (66.03%), indicating that individuals do adjust their SAH in response to health shocks. Rows 1 and 3 report average partial effects of SAH for the respondents who remained healthy, that is, were not hospitalized in the base period and rows 2 and 4 report the average partial effects of SAH for the respondents who had a recent health shock. Except for drugs, the effects of worse SAH are larger for the respondents who experienced a health shock. The interpretation is that for these respondents the variation in SAH is more likely to represent a variation in actual health. These findings again support the hypothesis that SAH measures health rather than personality or reporting style.

[Insert Table 6]

To sum up, we find robust evidence that SAH does have health content and that it predicts future health. Most of this health content reflects the actual health history of individuals. The remaining health content may be driven by various, unobserved individual factors, but they are not inconsequential and cannot be explained by reporting style.

5.3.Heterogeneity in the impact of SAH

Results presented in Table 7 show that SAH predicts certain illnesses better than others. As a proportion of the mean, the effects of poor SAH are especially large for cancer and diseases of the respiratory and endocrine systems. SAH is less predictive of skin and eye and ear diseases, which are arguably less serious. This evidence suggests that SAH responds more to symptoms associated with more serious illnesses. Although not reported, we find that, except for musculoskeletal disorders, SAH affects both in-hospital and all of the out-of-hospital services. For musculoskeletal disorders, poor SAH has large positive and statistically significant effects on drugs and certain out-of-hospital medical services (by rheumatologists, physiotherapists and podiatrists) but has no effect on in-hospital services.

[Insert Table 7]

Given the lack of evidence on the relationship between SAH and mental health in the literature (Jylhä, 2010), we provide more discussion on how SAH affects different health services related to mental disorders. As shown in Table 8, worse SAH has statistically and economically significant effects on hospitalizations and use of drugs for mental health problems. On the other hand, no significant effects of SAH are found on psychiatrist and psychologist visits. In contrast, we find that QoL does have large and statistically significant effects on the latter two variables as well as significantly affecting hospitalizations and drugs use. These findings suggest that SAH may not capture all mental health problems, and that QoL may be a more relevant measure in the analyses that focus on mental health.

[Insert Table 8]

Table 9 presents the average partial effects of SAH by gender and education (university education or not) and Figure 1 plots the average partial effects across the age distribution. Past literature has generally identified reporting heterogeneity in SAH along these dimensions so here we test whether such heterogeneity also carries through to the impact of SAH on future health. There is significant heterogeneity by age and gender, but the differences in the SAH effects by education are only (marginally) significant in the case of other medical services and prescription drugs. SAH is more predictive of future utilization for females than males which may be explained by the common trend that females are more likely to seek medical treatment than males. Meanwhile, we think that the lack of heterogeneous impact by education may be due to the counteracting effects of reporting bias and preference for health care. Past studies have found that highly educated individuals are more pessimistic when asked about their subjective health (Bago d' Uva et al., 2008) but at the same time they are also more likely to seek treatment when sick. The impact of SAH on future utilization varies non-linearly with age, especially for out-of-hospital services. The effects of SAH peak at around 55-60 years of age and decrease after. The decline is steeper in the case of poor SAH. The observed decline in the effects of SAH with age may be explained by fact that the expected health gains from an additional treatment for older individuals are smaller compared to an additional treatment for younger individuals, or alternatively that extra treatments may increase health risks. This result could also mean that SAH has less health content among older individuals.

[Insert Table 9]

[Insert Figure 1]

6. Conclusion

Health is complex and the notion of a perfect health index is difficult to define let alone measure. In this paper, we use a unique dataset constructed from merged survey and health administrative data sets to investigate whether the commonly used self-assessed health measure is in fact capturing "health" in a meaningful sense. Our empirical strategy consists of evaluating the predictive power of SAH on disaggregated measures of future health care utilization controlling for an extensive set of personal characteristics. We then use the utilization information to define comprehensive illness groups to investigate if SAH predicts some illnesses better than others. Additionally, we explore if there is variation in SAH effects by common sources of reporting heterogeneity in SAH responses (gender, age and education).

Our results suggest that the common self-reported health index does capture health in that it predicts future health service utilization, especially specialist visits and hospital admissions, over and above personal characteristics. The source of this predictive power is primarily actual health and the remaining sources are non-trivial health-related factors, ruling out reporting norms. Importantly, we find significant impact of SAH across all illness groups, but individuals appear to place more weight on symptoms leading to more serious illnesses, which SAH predicts better than less serious illnesses. The impact of SAH is found to be larger for females and younger (45+) individuals. Overall, our results confirm the interpretation of SAH as a useful indicator of objective health.

References

Arber, S., & Cooper, H. (2006). Gender differences in health in later life: the new paradox? Social Science and Medicine, 48(1), 61-76.

Bago d'Uva, T., Van Doorslaer, E., Lindeboom, M., & O'Donnell, O. (2008). Does reporting heterogeneity bias the measurement of health disparities? Health Economics, 17(3), 351-375.

Bound, J. (1991). Self-reported versus objective measures of health in retirement models. Journal of Human Resources, 26, 106-138.

Connelly, J.E., Philbrick, J.T., Smith, R.G., Kaiser, D.L., Wymer, A. (1989). Health perceptions of primary care patients and the influence on health care utilisation. Medical Care 27, S99–S109.

Crossley, T. F., & Kennedy, S. (2002). The reliability of self-assessed health status. Journal of Health Economics, 21(4), 643-658.

Ellis R.P., Fiebig D.G., Johar, M., Jones, G. & Savage, E. (2013). Explaining health care expenditure variation: large-sample evidence using linked survey and health administrative data. Health Economics, 22(9), 1093-1110.

Etilé, F., & Milcent, C. (2006). Income-related reporting heterogeneity in self-assessed health: evidence from France. Health Economics, 15(9), 965-981.

Gerdtham, U.-G., Johannesson, M., Lundberg, L., Isacson, D. (1999). A note on validating Wagstaff and van Dooslaer’s health measure in the analysis of inequality in health. Journal of Health Economics 18, 117-124.

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

Gupta, N. D., & Jürges, H. (2012). Do workers underreport morbidity? The accuracy of self-reports of chronic conditions. Social Science and Medicine 75, 1589-1594.

Hernandez-Quevedo, C., Jones, A.M. & Rice, N. (2004). Reporting bias and heterogeneity in self-assessed health. Evidence from the British Household Panel Survey. HEDG Working Paper 05/04, University of York.

Huisman, M., van Lenthe, F., & Mackenbach, J. (2007). The predictive ability of self-assessed health for mortality in different educational groups. International Journal of Epidemiology, 36(6), 1207-1213.

Humphries, K.H. & van Doorslaer, E. (2000). Income-related health inequality in Canada. Social Science and Medicine, 50, 663-671.

Idler, E.L. & Benyamini, Y. (1997). The reliability of self-assessed health status. Journal of Health and Social Behavior 38, 21-37.

Johar, M., Jones, G. & Savage, E. (2013). The effect of lifestyle choices on emergency department use in Australia. Health Policy 110(2-3): 280-290.

Johnston, D. W., Propper, C., & Shields, M. A. (2009). Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. Journal of Health Economics, 28(3), 540-552.

Jürges, H. (2007). True health vs response styles: exploring cross-country differences in self-reported health. Health Economics, 16(2), 163-178.

Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. Social Science andf Medicine, 69(3), 307-316.

Jylhä, M. (2010). Self-rated health between psychology and biology. A response to Huisman and Deeg. Social Science and Medicine, 70(5), 655-657.

Layes, A., Asada, Y., & Kephart, G. (2012). Whiners and deniers–What does self-rated health measure? Social Science and Medicine, 75(1), 1-9.

Lee (2000) The predictive value of self assessed general, physical, and mental health on functional decline and mortality in older adults. Journal of Epidemiology and Community Health, 54(2), 123-129.

Lindeboom, M., & Van Doorslaer, E. (2004). Cut-point shift and index shift in self-reported health. Journal of Health Economics, 23(6), 1083-1099.

Long, M. J., & Marshall, B. S. (1999). The relationship between self-assessed health status, mortality, service use, and cost in a managed care setting. Health Care Management Review, 24(4), 20-27.

Manor et al (2001) Manor, O., Matthews, S., & Power, C. (2001). Self-rated health and limiting longstanding illness: inter-relationships with morbidity in early adulthood. International journal of epidemiology, 30(3), 600-607.

McCallum, J., Shadbolt, B.,Wang, D., (1994). Self-rated health and survival: 7 years follow-up study of Australian elderly. American Journal of Public Health 847, 1100–1105.

Miilunpalo, S., Vuori, I., Oja, P., Pasanen, M., & Urponen, H. (1997). Self-rated health status as a health measure: the predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. Journal of Clinical Epidemiology, 50(5), 517-528.

Møller et al. (1996) Manor, O., Matthews, S., & Power, C. (2001). Self-rated health and limiting longstanding illness: inter-relationships with morbidity in early adulthood. International journal of epidemiology, 30(3), 600-607.

Salomon, J. A., Tandon, A., & Murray, C. J. (2004). Comparability of self-rated health: cross sectional multi-country survey using anchoring vignettes. British Medical Journal, 328(7434), 258.

van Doorslaer, E. and Gertham, U.-G. (2003), Does inequality in is self-assessed health predict inequality in survival by income? Social Science and Medicine 57, 1621-1629.

Table 1. Definitions of illness groups.

No	Hospital Diagnoses	Outpatient Specialities	Drug Groups
1	Infectious disease		Anti-infectives for systemic use
2	Neoplasm (cancer)	Medical Oncology	Antineoplastic and immuno-modulating agents
3	Blood disease	Haematology	Blood and blood forming organs
4	Endocrine, nutritional or metabolic disease	Endocrinology Immunology	Drugs for diabetes Systemic hormonal preparations
5	Mental disorder	Psychiatry Psychology	Mental disorders
6	Disease of nervous system	Neurology	Nervous system
7	Disease of eye	Ophthalmology	Ophthalmologicals
7	Disease of ear, nose or throat	Otorhinolaryngology	Otologicals
8	Disease of circulatory system	Cardiology	Cardiovascular system Lipid modifying agents
9	Disease of respiratory system	Thoracic medicine	Respiratory system
10	Disease of digestive system	Gastroenterology General surgery	Alimentary tract and metabolism
11	Disease of skin	Dermatology Plastic	Dermatological
12	Disease of musculoskeletal system	Rheumatology Orthopaedics Physiotherapy Podiatry	Musculoskeletal system
13	Genitourinary disease	Renal Urology Obstetrics and Gynaecology	Genitourinary system and sex hormones

Table 2. Means and sample proportions of main control variables by self-assessed health status

	Excellent/ very good	Good	Fair/poor
Quality of life good [a]	0.065	0.564	0.359
Quality of life fair or poor	0.010	0.067	0.517
Quality of life missing	0.021	0.027	0.037
Male	0.436	0.488	0.494
Age in years	60.86 (10.10)	63.65 (11.36)	65.89 (12.24)
Born in Australia [a]	0.765	0.754	0.747
Born in English speaking country	0.139	0.117	0.097
Speaks other language at home	0.072	0.104	0.134
Australian ancestry	0.521	0.524	0.529
English/Irish/Scottish ancestry	0.608	0.582	0.553
Other European ancestry	0.111	0.118	0.121
Other ancestry	0.127	0.147	0.168
Married/lives with partner	0.799	0.743	0.656
Number of children	2.41 (1.39)	2.47 (1.48)	2.53 (1.65)
Doesn't have any qualifications [a]	0.076	0.128	0.218
Has school/intermediate certificate	0.203	0.242	0.245
Has higher school certificate	0.099	0.101	0.097
Has trade/apprenticeship	0.098	0.127	0.129
Has certificate/diploma	0.229	0.210	0.171
HH income less than $5000 pa [a]	0.010	0.016	0.032
HH income $5000-$9999 pa	0.022	0.041	0.082
HH income $10000-$19999 pa	0.094	0.165	0.254
HH income $20000-$29999 pa	0.084	0.109	0.113
HH income $30000-$39999 pa	0.083	0.086	0.067
HH income $40000-$49999 pa	0.080	0.076	0.054
HH income $50000-$69999 pa	0.123	0.104	0.065
HH income missing	0.182	0.202	0.222
SEIFA Index of Relative Socioeconomic Advantage & Disadvantage	1014.15(87.85)	1001.34(82.94)	991.35(78.80)
Employed	0.569	0.437	0.255
Lives in a flat [a]	0.095	0.111	0.139
Lives in a house on farm	0.087	0.070	0.055
Lives in other housing	0.029	0.050	0.080
Accessibility/Remoteness Index of Australia [b]	1.22 (1.67)	1.25 (1.71)	1.26 (1.78)
Private health insurance	0.731	0.628	0.470
Diagnosed with skin cancer	0.261	0.262	0.259
Diagnosed with melanoma	0.051	0.058	0.069
Diagnosed with other cancer	0.047	0.071	0.111
Diagnosed with heart disease	0.068	0.145	0.255
Diagnosed with stroke	0.013	0.034	0.084
Diagnosed with diabetes	0.044	0.112	0.209
Diagnosed with blood clot	0.031	0.050	0.092

Diagnosed with asthma	0.097	0.128	0.185
Diagnosed with hay fever	0.153	0.156	0.157
Diagnosed with Parkinson's disease	0.003	0.006	0.018
Diagnosed with depression	0.111	0.164	0.282
Diagnosed with anxiety	0.071	0.107	0.182
Takes vitamins/supplements	0.516	0.512	0.470
Number of alcoholic drinks per week	7.24 (8.84)	7.05 (10.12)	6.09 (10.97)
Smokes now [a]	0.047	0.082	0.122
Smoked before, not now	0.336	0.371	0.396
Body mass index	26.16 (4.56)	27.88 (5.44)	28.93 (6.71)
Kessler psychological distress scale [c]	12.54 (3.58)	14.26 (4.90)	18.18 (7.41)
Long-term illness/disability	0.010	0.043	0.257
Physical Functioning scale [d]	92.37 (13.64)	79.54 (22.34)	51.31 (30.43)
Sample proportion	0.507	0.326	0.135

Notes: Sample size is 212,574.For continuous variables, standard deviations are in parentheses. [a] Omitted categories are excellent/very good quality of life, born in non-English speaking country, university degree, HH income $70,000 pa or more, lives in a house, and never smoked, respectively. [b] Varies from 0 (very accessible) to 15 (very remote). [c] Varies from 0 (no distress) to 50 (high distress). [d] Varies from 0 (low activity level) to 100 (high activity level).

Table 3. Means (sample proportions) of outcome variables by self-assesses health status

	Excellent/very good	Good	Fair/poor
HC service utilization:			
Hospital	0.202	0.277	0.394
GP	0.341	0.520	0.684
Specialist	0.419	0.525	0.647
Other medical	0.388	0.530	0.668
Drugs	0.264	0.486	0.710
Illnesses:			
Infection	0.215	0.346	0.531
Cancer	0.057	0.087	0.138
Blood	0.140	0.248	0.402
Endocrine	0.097	0.207	0.387
Mental	0.107	0.199	0.386
Nervous	0.132	0.268	0.490
Eye	0.513	0.560	0.594
Ear, nose, throat	0.057	0.079	0.116
Circulatory	0.378	0.577	0.724
Respiratory	0.098	0.173	0.308
Digestive	0.281	0.422	0.582
Skin	0.175	0.209	0.268
Musculoskeletal	0.213	0.346	0.492
Genitourinary	0.129	0.173	0.238
Sample proportion	0.507	0.326	0.135

Note: Sample size is 212,574. Details on the definition and the construction of the outcome variables are provided in the main text and Table 1.

Table 4. Average partial effects of SAH on future HC utilization

	Hospital	*GP*	*Specialist*	*Other medical*	*Drugs*
A. Quality of life variables not included					
Good SAH	0.033***	0.067***	0.057***	0.062***	0.065***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Fair/poor SAH	0.071***	0.101***	0.106***	0.106***	0.109***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Pseudo R2	0.069	0.164	0.082	0.127	0.394
B. Quality of life variables included					
Good SAH	0.035***	0.067***	0.060***	0.063***	0.067***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Fair/poor SAH	0.073***	0.106***	0.113***	0.114***	0.115***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)
Good QoL	-0.004	-0.000	-0.005	-0.002	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Fair/poor QoL	-0.002	-0.013**	-0.016**	-0.019***	-0.013***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)
Pseudo R2	0.069	0.164	0.082	0.127	0.394
%Change: good SAH	6.2%	0.6%	4.9%	2.3%	2.7%
%Change: fair/poor SAH	3.0%	4.7%	7.0%	7.3%	5.4%
Mean of dependent var.	0.256	0.453	0.487	0.478	0.406

Notes: Sample size is 212,574. Robust standard errors are in parentheses. All logit regressions control for demographic and socio-economic characteristics, health behaviors, self-reported health measures, and family health history. The models in panel B additionally control for quality of life. The omitted group for SAH and QoL is excellent or very good. The bottom panel reports percentage change in the estimated average partial effects of SAH when quality of life variables are added. Symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 5. Average partial effects of SAH on future HC utilization controlling for past utilization

	Hospital	*GP*	*Specialist*	*Other medical*	*Drugs*
A. Baseline model					
Good SAH	0.035***	0.067***	0.060***	0.063***	0.067***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Fair/poor SAH	0.073***	0.106***	0.113***	0.114***	0.115***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)
Pseudo R2	0.069	0.164	0.082	0.127	0.394
B. Past utilization added					
Good SAH	0.015***	0.029***	0.018***	0.023***	0.011***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Fair/poor SAH	0.031***	0.039***	0.033***	0.038***	0.018***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Pseudo R2	0.119	0.249	0.220	0.229	0.636
C. Past utilization disaggregated (into quarters)					
Good SAH	0.015***	0.023***	0.014***	0.020***	0.008***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
Fair/poor SAH	0.029***	0.028***	0.025***	0.031***	0.014***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Pseudo R2	0.124	0.296	0.237	0.270	0.651
%Change: good SAH	-57.7%	-66.2%	-76.6%	-69.1%	-87.7%
%Change: fair/poor SAH	-60.4%	-73.6%	-77.8%	-73.0%	-87.9%
Mean of dependent var.	0.256	0.453	0.487	0.478	0.406

Note: Sample size is 212,574. Robust standard errors are in parentheses. All logit regressions control for demographic and socio-economic characteristics, health behaviors, self-reported health measures, family health history, and quality of life. The omitted group for SAH is excellent or very good. The bottom panel reports percentage change in the estimated average partial effects of SAH from the baseline when disaggregated past utilization measures are added. Symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 6. Variation in average partial effects of SAH in "healthy" subsample

	Hospital	GP	Specialist	Other medical	Drugs
Good SAH					
Health shock: no	0.009	0.030***	0.019**	0.032***	0.010*
	(0.005)	(0.006)	(0.007)	(0.007)	(0.004)
Health shock: yes	0.007	0.031*	0.027	0.035*	0.034***
	(0.012)	(0.013)	(0.014)	(0.014)	(0.008)
Fair/poor SAH					
Health shock: no	0.019	0.022	0.035*	0.035**	0.016*
	(0.011)	(0.012)	(0.014)	(0.013)	(0.007)
Health shock: yes	0.056*	0.064**	0.124***	0.140***	0.020
	(0.024)	(0.024)	(0.030)	(0.029)	(0.013)

Note: Sample size is 37,243. Robust standard errors are in parentheses. Health shock is defined as a hospitalization in the past 12 months. All logit regressions control for demographic and socio-economic characteristics, health behaviors, self-reported health measures, family health history, quality of life, and health care utilization in the past year. The omitted group for SAH is excellent or very good. Symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 7. Variation in average partial effects of SAH by illnesses

	Good SAH			Fair/poor SAH		
	Coeff.	S.e.	% change from mean	Coeff.	S.e.	% change from mean
Cancer	0.028***	(0.002)	35.8	0.064***	(0.004)	81.0
Respiratory	0.037***	(0.002)	24.1	0.085***	(0.004)	55.1
Endocrine	0.041***	(0.002)	23.1	0.090***	(0.004)	51.1
Blood	0.038***	(0.002)	17.4	0.078***	(0.004)	36.1
Digestive	0.072***	(0.003)	19.3	0.132***	(0.005)	35.3
Nervous	0.041***	(0.002)	17.8	0.071***	(0.004)	30.8
Genitourinary	0.025***	(0.002)	15.5	0.049***	(0.004)	30.7
Ear, nose, throat	0.009***	(0.002)	11.8	0.021***	(0.003)	29.1
Mental	0.021***	(0.002)	11.5	0.046***	(0.003)	25.8
Circulatory	0.080***	(0.003)	16.1	0.123***	(0.004)	24.8
Infection	0.032***	(0.003)	10.4	0.069***	(0.004)	22.6
Musculoskeletal	0.041***	(0.003)	13.8	0.056***	(0.004)	18.7
Skin	0.007**	(0.002)	3.3	0.020***	(0.004)	10.1
Eye	0.018***	(0.003)	3.3	0.027***	(0.005)	4.9

Note: Sample size is 212,574. All logit regressions control for demographic and socio-economic characteristics, health behaviors, self-reported health measures, family health history, and quality of life. S.e. denotes robust standard errors. The omitted group for SAH is excellent or very good. Symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 8. Average partial effects of SAH and QoL on mental health

	Hospital	Psychiatrist	Psychologist	Drugs
Good SAH	0.004***	0.001	-0.001	0.024***
	(0.001)	(0.001)	(0.001)	(0.002)
Fair/poor SAH	0.007***	0.002*	-0.001	0.045***
	(0.001)	(0.001)	(0.001)	(0.003)
Good QoL	0.002*	0.003***	0.005***	0.009***
	(0.001)	(0.001)	(0.001)	(0.002)
Fair/poor QoL	0.003**	0.004***	0.009***	0.009**
	(0.001)	(0.001)	(0.002)	(0.003)
Mean of dependent var.	0.010	0.012	0.015	0.165
Pseudo R2	0.172	0.231	0.144	0.263

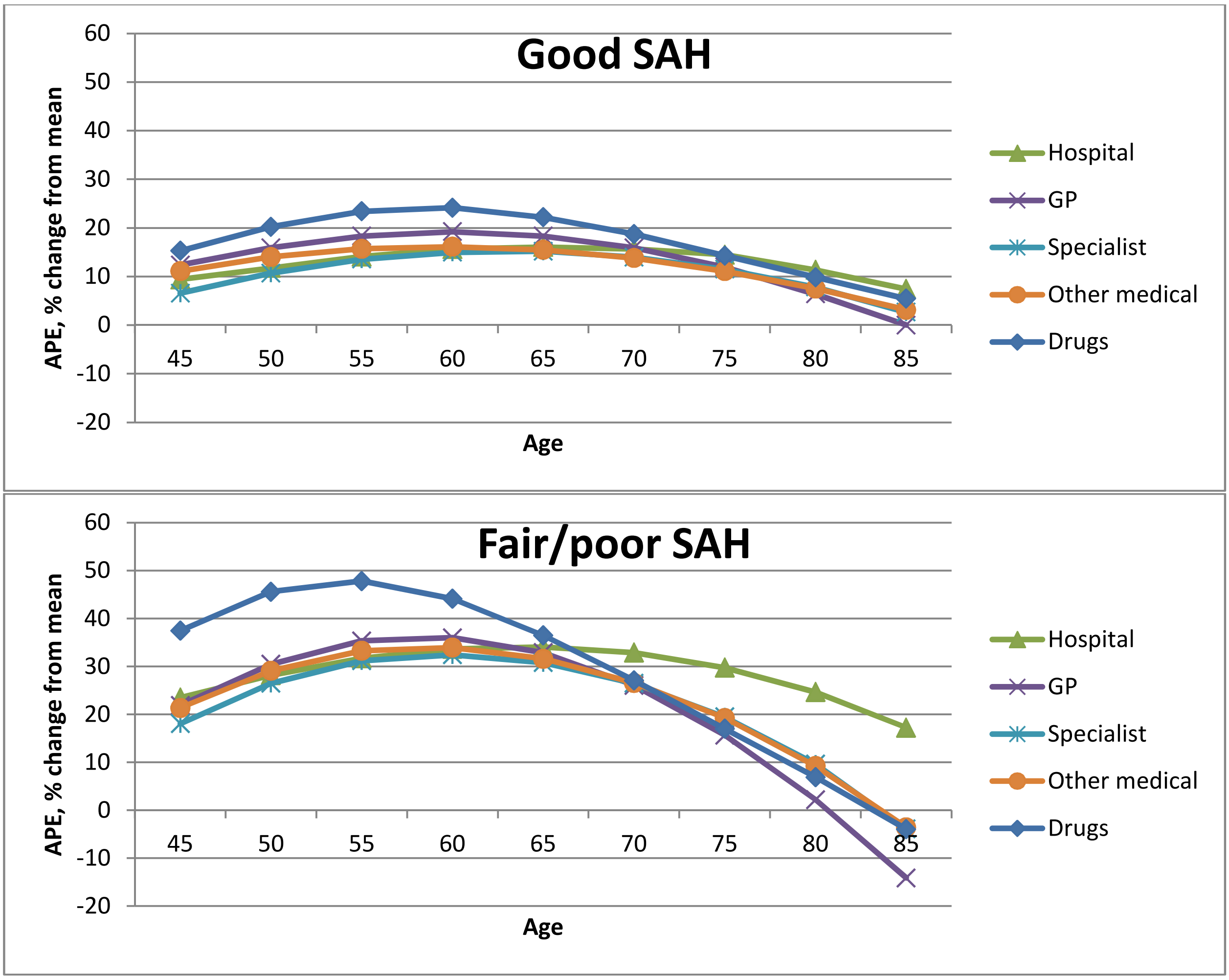
Note: Sample size is 212,574. Robust standard errors are in parentheses. All logit regressions control for demographic and socio-economic characteristics, health behaviors, self-reported health measures, family health history, and quality of life. The omitted group for SAH and QoL is excellent or very good. Symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 9. Heterogeneity in average partial effects of SAH by gender and education, ppt

	Hospital	GP	Specialist	Other medical	Drugs
A. Good SAH					
Female	0.032***	0.075***	0.056***	0.067***	0.072***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)
Male	0.036***	0.055***	0.061***	0.056***	0.061***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
No university degree	0.034***	0.066***	0.061***	0.062***	0.068***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
University degree	0.036***	0.069***	0.049***	0.062***	0.068***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
B. Fair/poor SAH					
Female	0.070***	0.133***	0.125***	0.137***	0.133***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)
Male	0.081***	0.093***	0.116***	0.105***	0.105***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)
No university degree	0.074***	0.114***	0.122***	0.119***	0.119***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
University degree	0.078***	0.118***	0.118***	0.138***	0.130***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)
Pseudo R2	0.070	0.166	0.084	0.128	0.396
Mean of dependent var.	0.256	0.453	0.487	0.478	0.406

Notes: Sample size is 212,574. Robust standard errors are in parentheses. All logit regressions control for demographic and socio-economic characteristics, self-reported health behaviors, health measures, family health history, and quality of life. The omitted group for SAH is excellent or very good. Symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Figure 1. Heterogeneity in average partial effects of SAH by age.



Notes: Sample size is 212,574. All logit regressions control for demographic and socio-economic characteristics, self-reported health behaviors, health measures, family health history, and quality of life. The omitted group for SAH is excellent or very good.