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 reporting biases and interpretation, 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. Our large sample size allows very disaggregate analysis and we find that SAH predicts more serious, chronic illnesses better than less serious illnesses. Finally we compare the predictive power of SAH relative to administrative data and an extensive set of self-reported health measures, SAH does not add to the predictive power of future utilization when the administrative data is included and improves prediction only marginally when the extensive survey-based health measures are included. Clearly there is value in the more extensive survey and administrative health data as well as greater cost of collection.

Running title: Does self-assessed health measure health?

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

JEL codes: I1

[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. We thank a referee for useful and constructive comments.

* 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

1. Introduction

Accounting for variations in health has become standard in many applied fields outside health economics. Prime examples include the importance of health in labour (Bound, 1991), consumption and savings (Jones and O'Donnell, 1995; De Nardi et al., 2010), portfolio choices (Rosen and Wu, 2004) and tourism (Hunter-Jones and Blackburn, 2007). Recent results by Finkelstein, Luttmer and Notowidigdo (2013) suggest that variations in health impact on fundamentals in complex ways such as changing the curvature of the utility function, further supporting the inclusion of health measures in various fields of study.

The notion of "true" health is itself a complex and multi-dimensional concept and various different measures of health have been used in practice depending on the availability of data and the research question. Still, the most pervasive measure remains self-assessed health (SAH). This variable is based on a survey question asking respondents to rate their general health usually on a five-point scale. It is easily asked and hence more readily available to researchers than most other measures. 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).

Despite its popularity, concerns remain about the validity of the SAH variable (we provide more details with references below) and there is a lot of uncertainty regarding what exactly SAH is measuring. In this paper we ask two main questions: Does SAH measure health? How well does SAH perform compared to other more objective measures of health in predicting future health? We are able to address these questions with the use of a rich data set that combines administrative panel data on detailed and

comprehensive health care utilization with survey data also containing an extensive range of health questions along with the standard socio-economic and demographic characteristics. To our knowledge, we are the first paper to assess the usefulness of SAH using a large sample survey linked to comprehensive health claims data.

The empirical strategy involves the use of prospective models to investigate the association between SAH and a range of future health outcomes as measured by future health care utilization. We look separately at the relationship of SAH with various types of health care (hospital admissions, out-of-hospital medical services and prescription drugs) as well as major illnesses. We also compare the performance of SAH as a predictor of future health to that of other more objective survey health measures and to the administrative health data. Finally we briefly investigate the interaction of SAH with other non-health controls in predicting future health in the absence of extensive health controls. This may be useful to researchers faced with the usual dataset including a SAH variable but no other health measures and who have to make decisions on which controls to include in their specifications.

We find that low self-assessed health substantially increases the utilization of health care services. For example, in specifications with a wide range of socio-economic and demographic controls as well as family health history variables, the effects of poor SAH (relative to excellent) vary from 39 percent to 54 percent relative to the mean utilization rates across different health care services. Interestingly, the effects of SAH are stronger for the more serious illnesses such as cancer. We take these results as convincing evidence that SAH does have actual health content.

However, our results also suggest that the predictive power of SAH is surpassed by both the administrative data and by self-reported variables on health problems.

Specifically, objective health measures obtained from past administrative data (e.g., diagnoses from past hospitalization) are the most predictive of an individual's future health care utilization followed by self-reported health measures. Hence there is a value in investing in data collection, especially administrative records and including extensive health questions in a survey.

2. Literature on SAH

A major concern with the use of SAH as measuring "health" stems from results on reporting biases related to various individual characteristics. For example, researchers have found discrepancies between SAH and more objective health measures related to sex (Arber and Cooper, 2006), income (Humphries and van Doorslaer, 2000; Hernandez-Quevedo et al., 2004; Etilé and Milcent, 2006) and nationality (Jürges, 2007). An alternative approach based on vignette anchoring has suggested variation in reporting behaviour based on age (Lindeboom and van Doorslaer, 2004), education (Bago d'Uva et al., 2008) and country of residence (Salomon et al., 2004).

Crossley and Kennedy (2002) use a natural experiment where some respondents answer the SAH question twice to 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 (Johnston et al., 2009) and the presence of incentives for misreporting (Bound, 1991; Gupta and Jürges, 2012 and references within). Given the number and variety of reporting biases found in the literature, it is natural and appropriate to ask what SAH actually measures (Layes et al., 2012 and references therein).

Several studies have focused on the predictive power of SAH on survival. Early papers found that SAH did in fact predict mortality (McCallum et al., 1994 and studies reviewed in Idler and Benyamini, 1997) but tended to concentrate on smaller samples of older adults. More recent papers also found that SAH has predictive power for mortality and moreover that the predictive power varies negatively with age (van Doorslaer and Gerdtham, 2003) and positively with education at least for men (Huisman et al., 2007). Jylhä (2009) discusses pathways in which SAH may predict mortality.

Recent studies find that SAH has predictive power with respect to other health outcomes as well. Lee (2000) finds that SAH predicts not only mortality, but also functional decline while 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. Wu et al. (2013) find that unhealthy individuals have at least twice the odds of contracting an array of chronic diseases than their healthy counterparts, and are also more likely to have poor spiritual status and poor quality of interpersonal relationships. 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. Finally, Vie et al. (2014) find that SAH during adolescence (aged 12-20) is a significant predictor of allostatic load in adulthood.

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 linked to administrative emergency

department records, Johar et al. (2013) find a strong negative SAH gradient in presentations to an emergency department. Ellis et al. (2013) show that SAH has strong predictive power over aggregate health care expenditure, but did not explore which component specifically of the health care that SAH predicts.

The contribution of this study revolves around the superior dataset we use. As discussed above, existing studies have focused on restrictive samples and/or restrictive health outcomes. With a large representative sample of adults over 45 years of age we are able to look at the predictive power of SAH on a comprehensive set of future health outcomes in a consistent fashion. In particular, the large size of our sample makes it possible to identify and credibly examine relationships between SAH and relatively rare illnesses, such as cancer. Secondly, most of the above mentioned studies use self-reported illnesses and physical health scores, such as SF-36, as measures of objective health, and these may be reported with errors. This study is based on both extensive administrative and self-reported survey data. Hence we have access to more accurate data and we can compare the performance of health measures based on self-reports and those based on actual utilization of health services. 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. Finally, the richness of the survey data allows us to include extensive controls such as family history of chronic illnesses that have been omitted in the past.

3. Method

The aim of this study is first 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 along with SAH to understand the impact of confounding factors. Specifically, we estimate multivariate models of various health care utilization measures at time t+1 as functions of SAH and extensive control variables at time t. These models will confirm whether or not SAH has independent effects on prospective health outcomes. Also, 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. We exclude a set of these control variables one at a time to assess the sensitivity of SAH coefficients to omitted variables.

As an additional analysis, we restrict attention to a homogenous sub-sample of relatively healthy individuals and we compare the effect of SAH on future health care utilization for the sub-group of these healthy individuals who experience a negative health shock in the form of a hospital admission at time t-1 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 ago. 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.

We also 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 are estimated for each illness group.

Additionally, we explore heterogeneity in the health content of SAH along the dimensions of age, sex and education by including their interaction terms with SAH.

Finally, we assess the relative usefulness of SAH compared to other health measures by replacing SAH with self-reported health problems from the survey and past utilization from the administrative data. We record the explanatory power of these models (using pseudo R-squared) and compare them.

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 percent 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 percent). 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 includes prescription drugs for which a Medicare subsidy was paid during 2006-2009. About 80 percent 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 percent of observations had missing values, the observations were kept and dummy variables for missing information were added to the model. If less than 1 percent of observations had missing values, the observations with missing values were deleted. Our final analysis sample consists of 212,574 observations (80 percent 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-length general practitioner (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, ophthalmology, pathology, physiotherapy, podiatry and radiation oncology;
5. *Drugs*, a binary variable equal to 1 if the respondent consumed more than 2 types of drug groups (explained below).

For all 5 variables, the threshold used to switch the dummy variable to 1 corresponds to the sample median. We prefer this binary transformation because it facilitates comparison across types of utilization and because counts are sensitive to rare, extreme observations, which may lead to misleading results. Nevertheless, as a robustness check, we also estimate negative binomial models for count data. To avoid bias due to extreme users, for each outcome, we exclude observations in the top 1 percent of the distribution.

To define drug groups or types of drugs, we use the first digit of the Anatomical Therapeutic Codes (ATC). We focus on drug groups rather than individual drugs for several reasons. A consumption of multiple drug groups indicates comorbidities.

Also, the use of drug groups minimises potential bias due to the data limitation which only covers subsidized drugs; essentially we are exploiting substitutability or complementary between drugs for a given condition, where at least one of them is 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 percent), 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 percent excellent or very good, 32.6

percent good and 13.5 percent fair or poor (3.2 percent of the sample did not report their SAH).

We compare the predictive power of SAH to two types of more objective health measures. The first group comes from the 45 and Up survey and consists of self-reported diagnoses of illnesses (skin cancer, melanoma, other cancer, heart disease, stroke, diabetes, blood clot, asthma, hay fever, Parkinson's disease, depression and anxiety) and daily health limitations (physical functioning and mental distress, as measured by the Kessler psychological distress scale), 15 variables in total. The second group of objective health measures includes past utilization variables that are constructed using the historical dimension of the linked administrative data. In effect, we are including a vector of lagged dependent variables, although all past utilization measures are disaggregated by illness and the time frame covered by these lagged variables is in some cases longer than the one-year-ahead time frame used for the dependent variables. Specifically, we add illness-specific hospital admissions in the past five years and out-patient service and prescription drug use in the last 12 months from the survey date (136 variables).

We also control for an extensive list of individual characteristics. The survey data contain 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), as well as family health history (parents' and siblings' illnesses). We also control for self-reported quality of life (QoL). Like SAH, QoL is likely to be influenced by reporting norms and biases and its inclusion in the model will help control for the reporting style of respondents.

Although some of the control variables may be correlated with unobserved aspects of health, we refer to them as "non-health" variables. In total, there are 65 such controls (98 variables).

5. Results

5.1. Descriptive statistics

Table 2 presents the means of the main control variables by self-assessed health status. People in worse SAH are older and/or more likely to 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. As expected there is a strong and positive correlation between SAH and quality of life.

[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 illnesses in the next 12 months is also higher among the respondents in worse health.

[Insert Table 3]

5.2.Health content of SAH

Our first research question asks how much health content there is in the self-assessed health variable. Panel A of Table 4 presents the logit average partial effects of good and poor SAH (relative to excellent or very good health) on the utilization of various health services in the next 12 months. The regressions in panel A control for all non-health variables described in Section 4. The results show that there is a significant positive relationship between worse SAH and future health care utilization even after controlling for extensive demographic, socio-economic, lifestyle, family health history and quality of life variables. 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 5 percentage points higher probability of hospitalization and 9-10 percentage points higher probability of out-of-hospital services utilization, while those in fair/poor health have 14 percentage points higher probability of hospitalization and 18-21 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. Relative to the mean utilization rates, the effects of poor SAH are largest on hospitalizations (54 percent relative to the mean of 0.256) and prescription drug use (52 percent relative to the mean of 0.406).

[Insert Table 4]

We also check how important it is to control for the non-health variables, especially the variables that may not be available in other surveys. Panels B, C, D, E and F of Table 4 show how the average partial effects of good and poor SAH change when we omit demographic, socio-economic status (SES), lifestyle, family health history and quality of life (QoL) variables, respectively. The coefficients on SAH are somewhat affected by the omission of SES, lifestyle and QoL variables. For example, the effect of poor SAH on GP visits increases from 18.1 to close to 20 percentage points when SES, lifestyle or QoL variables are not included in the regression. The effects of SAH on the use of other medical services and prescription drugs also increase slightly, but SAH effects on specialist visits and hospitalizations are quite insensitive to the omission of these controls. Omitting demographic variables or information about family health history affects the coefficients on the SAH variables even less than omitting SES, lifestyle or QoL variables. These findings suggest that it may not be necessary to have such a rich set of controls as we do in our models. A more parsimonious model, which controls for the usual individual characteristics and QoL (or other variables capturing an individual's reporting style) may suffice.

As a sensitivity check on the decision to discretise the health outcomes into binary variables, we estimate negative binomial models for the number of GP and specialist visits, other medical services and prescription drug types. We do not perform this exercise for hospitalization because most hospitalization cases happen only once in a year. To avoid excessive influence by outliers, we omit the one percent highest users of each health care service (in the estimation of the regression on that service only). Two models are estimated. The first is a standard count data model, while the second is a latent class model with two classes. The latter model allows for unobserved heterogeneity in the effects of SAH and other variables. (For drugs, the latent class

model has convergence problems, so we only present negative binomial model results.)

The results of the count data models, presented in Table 5, are consistent with the results of the logit models. For all dependent variables and both types of count data 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.

[Insert Table 5]

Table 6 presents results for the subset of respondents classified as "healthy". Close to 18 percent of the sample are defined as being relatively healthy according to our definition (described in Section 3). As expected, the proportion of individuals in excellent or very good health is higher in this subsample (65.36 percent) than in the rest of the sample (48.37 percent). Around 12 percent of the "healthy" individuals had a health shock (a hospitalization) in the past year. The proportion of individuals in excellent or very good SAH is lower among the hospitalized respondents (60.43 percent) than among those who did not have a hospitalization (66.03 percent), 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 past 12 months and rows 2 and 4 report the average partial effects of SAH for the respondents who had a recent health shock.[2] The effects of worse SAH on all outcomes are larger for the respondents who experienced a health shock. The effect of poor SAH on GP visits is twice as large in the subsample of people who had a hospitalization last year as in the subsample of

[2] In rows 1 and 3, the reference group is people with excellent health who have not experienced a health shock. In rows 2 and 4, the reference group is people with excellent health who have experienced a health shock. The effect of excellent health is normalized to zero for both groups.

people who were not hospitalized. The differences in the effects of poor SAH between these two groups are even larger in the case of specialist visits, other medical service use and hospitalizations. The interpretation of these results 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]

5.3.Heterogeneity in the impact of SAH

Results presented in Table 7 show that SAH is more closely related to certain illnesses than others. As a proportion of the mean, poor SAH has especially large effects on the probabilities of cancer and diseases of the respiratory and endocrine systems. SAH is least related to skin and eye diseases, which are arguably less serious. This evidence suggests that SAH responds more to symptoms associated with more serious illnesses and is consistent with the results of Manor et al. (2001) on British data. 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]

Table 8 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 interestingly the differences in SAH effects by education are not statistically significant. SAH is more strongly related to future health care 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 than in the case of good SAH. The observed decline in the effects of SAH with age may be explained by the 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 8]

[Insert Figure 1]

5.4.Comparison of the predictive power of SAH and more objective health measures

Our second research question looks at the predictive power of SAH relative to other health measures. To address this issue, we estimate models with various combinations of the three groups of health variables (SAH, the self-reported diagnoses of illnesses obtained from the survey data and the past health care utilization variables obtained from the administrative records) and compare the pseudo R-squared statistics from these regressions. Table 9 presents the pseudo-R-squared statistics for the five future health care utilization measures we are predicting. The first row presents the statistics for models without any health controls. The rows below it present corresponding figures for models in which the three groups of health variables are included in various combinations.

[Insert Table 9]

The use of our extensive set of survey-based objective health variables instead of the one SAH variable (comparing rows 2 and 3) raises the pseudo R-squared statistic by 7 to 25 percent across the five models. When comparing SAH to the information from the administrative data (comparing rows 2 and 4), the pseudo R-squared rises by a factor of up to 3. When adding SAH to regressions containing the other survey-based health measures (comparing rows 3 and 5), the pseudo R-squared increases by 1 to 3 percent only while the impact of adding SAH to the administrative data (comparing rows 4 and 6) is negligible.

Our results show that, given the choice, one would prefer the use of the administrative data. This is perhaps not surprising since in this case, the administrative variables are in essence lagged values of the dependent variables. The more interesting result is that although adding health information (the coefficients on SAH remain significant in all models), SAH does not improve the predictive power in the case of the administrative

data and does so only marginally in the case of the survey-based objective health measures. Having said this, we are comparing the power of one easily collected variable with a set of 15 possibly sensitive, survey-based health variables and with rare administrative utilization data (136 variables). The overall conclusion is that there is value in collecting the more extensive data and using them in predicting future health care utilization. However, in the absence of administrative information, both SAH and self-reports of illnesses should be used to predict future health.

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 usefulness of SAH compared to more objective health measures and its predictive power on various types of future health care utilization variables. We then use the utilization information to define comprehensive illness groups to investigate if SAH is more closely related to 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).

All of our results suggest that the common self-reported health index does capture health in that it is strongly related to future health service utilization, especially specialist visits and hospital admissions, over and above personal characteristics. 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, however, compared to objective health measures, SAH has less predictive power. Given the choice, inclusion of administrative data would perform better followed by survey-based illnesses measures.

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.

De Nardi, M., French E. & Jones, J.B. (2010). Why do the elderly save? The role of medical expenses. Journal of Political Economy, 118, 37-75.

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.

Finkelstein, A., Luttmer, E.F.P. & Notowidigdo, M.J. (2013). What good is wealth without health? The effect of health on the marginal utility of consumption. Journal of the European Economic Association, 11(S1), 221-258.

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, United Kingdom.

Hunter‐Jones, P. & Blackburn, A. (2007). Understanding the relationship between holiday taking and self‐assessed health: an exploratory study of senior tourism. International Journal of Consumer Studies, 31(5), 509-516.

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.

Jones, A. & O'Donnell, O. (1995). Equivalence scales and the costs of disability. Journal of Public Economics, 57, 273-289.

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 and Medicine, 69(3), 307-316.

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, Y. (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, 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, L., Kristensen, T.S & Hollnagel, H. (1996). Self-rated health as a predictor of coronary heart disease in Copenhagen, Denmark. Journal of Epidemiology and Community Health, 50, 423-428.

Rosen, H.S. & Wu, S. (2004). Portfolio choice and health status. Journal of Financial Economics, 72(3), 457-484.

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.

Vie, T.L, Hufthammer, K.O., Holmen, T.L., Meland, E. & Breidablik, H.J. (2014). Is self-rated health a stable and predictive factor for allostatic load in early adulthood? Findings from the Nord Trøndelag Health Study (HUNT). Social Science and Medicine, 117, 1-9.

Wu, S., Wang, R., Zhao, Y., Ma, X., Wu, M., Yan, X. & He, J. (2013). The relationship between self-rated health and objective health status: a population-based study. BMC Public Health, 13(1), 320.

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
8	Disease of ear, nose or throat	Otorhinolaryngology	Otologicals
9	Disease of circulatory system	Cardiology	Cardiovascular system Lipid modifying agents
10	Disease of respiratory system	Thoracic medicine	Respiratory system
11	Disease of digestive system	Gastroenterology General surgery	Alimentary tract and metabolism
12	Disease of skin	Dermatology Plastic	Dermatological
13	Disease of musculoskeletal system	Rheumatology Orthopaedics Physiotherapy Podiatry	Musculoskeletal system
14	Genitourinary disease	Renal Urology Obstetrics and Gynaecology	Genitourinary system and sex hormones

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

	Excellent/ very good	Good	Fair/poor
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

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)
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
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 born in non-English speaking country, university degree, HH income $70,000 pa or more, lives in a house, and never smoked, excellent/very good quality of life, respectively. [b] Varies from 0 (very accessible) to 15 (very remote).

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

	Excellent/very good	Good	Fair/poor
HC service utilization:			
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
Hospital	0.202	0.277	0.394
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 health care utilization across different specifications

	GP	*Specialist*	*Other medical services*	*Drugs*	*Hospital*
A. All non-health controls					
Good SAH	0.094***	0.087***	0.091***	0.102***	0.053***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Fair/poor SAH	0.181***	0.192***	0.188***	0.209***	0.137***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
B. Demographics left out					
Good SAH	0.102***	0.090***	0.095***	0.110***	0.059***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Fair/poor SAH	0.187***	0.192***	0.190***	0.214***	0.145***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
C. SES left out					
Good SAH	0.102***	0.087***	0.095***	0.112***	0.053***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Fair/poor SAH	0.199***	0.192***	0.197***	0.244***	0.136***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
D. Lifestyle left out					
Good SAH	0.103***	0.089***	0.096***	0.115***	0.056***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Fair/poor SAH	0.197***	0.195***	0.195***	0.233***	0.144***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
E. Family health history left out					
Good SAH	0.098***	0.090***	0.094***	0.107***	0.054***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Fair/poor SAH	0.189***	0.197***	0.194***	0.219***	0.140***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
F. QoL left out					
Good SAH	0.101***	0.089***	0.094***	0.106***	0.055***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Fair/poor SAH	0.199***	0.201***	0.194***	0.224***	0.155***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mean of dep.var.	0.453	0.487	0.478	0.406	0.256

Notes: Sample size is 212,574. Average partial effects are based on logit models. Robust standard errors are in parentheses. Non-health controls include demographic, SES, lifestyle, family health history and QoL variables and year effects. The omitted group for SAH is excellent or very good. Symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 5. Count data model results.

	GP	*Specialist*	*Other medical services*	*Drugs*
A. Negative binomial model				
Good SAH	0.139***	0.253***	0.164***	0.251***
	(0.005)	(0.011)	(0.006)	(0.005)
Fair/poor SAH	0.219***	0.449***	0.292***	0.360***
	(0.007)	(0.016)	(0.009)	(0.008)
B. Latent class negative binomial model				
Latent class 1				
Good SAH	0.083	0.218***	0.266***	-
	(0.053)	(0.013)	(0.017)	-
Fair/poor SAH	0.266**	0.415***	0.426***	-
	(0.082)	(0.018)	(0.026)	-
Class 1 probability	0.065	0.334	0.339	-
Latent class 2				
Good SAH	0.192***	0.717***	0.202***	-
	(0.005)	(0.028)	(0.009)	-
Fair/poor SAH	0.366***	1.405***	0.479***	-
	(0.007)	(0.038)	(0.013)	-
Class 2 probability	0.935	0.666	0.661	
Sample size	210613	210651	210530	211686
Mean of dep.var.	7.57	2.11	14.45	2.66

Notes: Sample size is 212,574. Robust standard errors are in parentheses. All count regressions control for demographic, SES, lifestyle, family health history and QoL variables and year effects. The omitted group for SAH is excellent or very good. Symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

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

	GP	*Specialist*	*Other medical services*	*Drugs*	*Hospital*
Good health					
Health shock: No	0.034***	0.026***	0.038***	0.015***	0.012*
	(0.006)	(0.007)	(0.007)	(0.004)	(0.005)
Health shock: Yes	0.035**	0.039**	0.045**	0.040***	0.015
	(0.013)	(0.015)	(0.014)	(0.008)	(0.013)
Fair/poor health					
Health shock: No	0.038**	0.063***	0.057***	0.032***	0.029**
	(0.012)	(0.014)	(0.013)	(0.007)	(0.011)
Health shock: Yes	0.080***	0.171***	0.174***	0.042**	0.089***
	(0.024)	(0.030)	(0.029)	(0.013)	(0.025)

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, SES, lifestyle, family health history and QoL variables, year effects, 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
Respiratory	0.109***	(0.003)	70.88	0.194***	(0.004)	125.95
Cancer	0.033***	(0.002)	42.05	0.086***	(0.004)	108.98
Endocrine	0.077***	(0.002)	43.46	0.189***	(0.004)	106.72
Blood	0.064***	(0.002)	29.94	0.161***	(0.004)	74.65
Nervous	0.066***	(0.002)	28.41	0.160***	(0.004)	69.39
Mental	0.041***	(0.002)	23.09	0.112***	(0.004)	62.58
Digestive	0.093***	(0.003)	24.97	0.199***	(0.004)	53.32
Genitourinary	0.032***	(0.002)	20.11	0.076***	(0.004)	47.45
Musculoskeletal	0.066***	(0.003)	22.14	0.138***	(0.004)	45.95
Infection	0.052***	(0.003)	16.95	0.133***	(0.004)	43.42
Ear, nose, throat	0.012***	(0.002)	16.22	0.031***	(0.003)	43.25
Circulatory	0.065***	(0.002)	13.09	0.167***	(0.004)	33.65
Skin	0.013***	(0.002)	6.69	0.045***	(0.004)	22.69
Eye	0.025***	(0.003)	4.63	0.045***	(0.004)	8.34

Note: Sample size is 212,574. S.E. denotes robust standard errors. All logit regressions control for demographic, SES, lifestyle, family health history and QoL variables and year effects. The omitted group for SAH is excellent or very good. Symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. [a] Mean refers to the mean of the dependent variable.

Table 8. Heterogeneity in average partial effects of SAH by gender and education

	GP	*Specialist*	*Other medical services*	*Drugs*	*Hospital*
A. Good SAH					
Female	0.104***	0.084***	0.094***	0.109***	0.053***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)
Male	0.081***	0.088***	0.084***	0.097***	0.052***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
No university degree	0.093***	0.089***	0.090***	0.103***	0.052***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
University degree	0.097***	0.076***	0.090***	0.107***	0.053***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
B. Fair/poor SAH					
Female	0.211***	0.202***	0.207***	0.236***	0.139***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Male	0.168***	0.191***	0.182***	0.206***	0.140***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
No university degree	0.193***	0.200***	0.193***	0.222***	0.141***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
University degree	0.192***	0.188***	0.206***	0.231***	0.134***
	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)
Pseudo R-squared	0.149	0.065	0.108	0.364	0.058
Mean	0.453	0.487	0.478	0.406	0.256

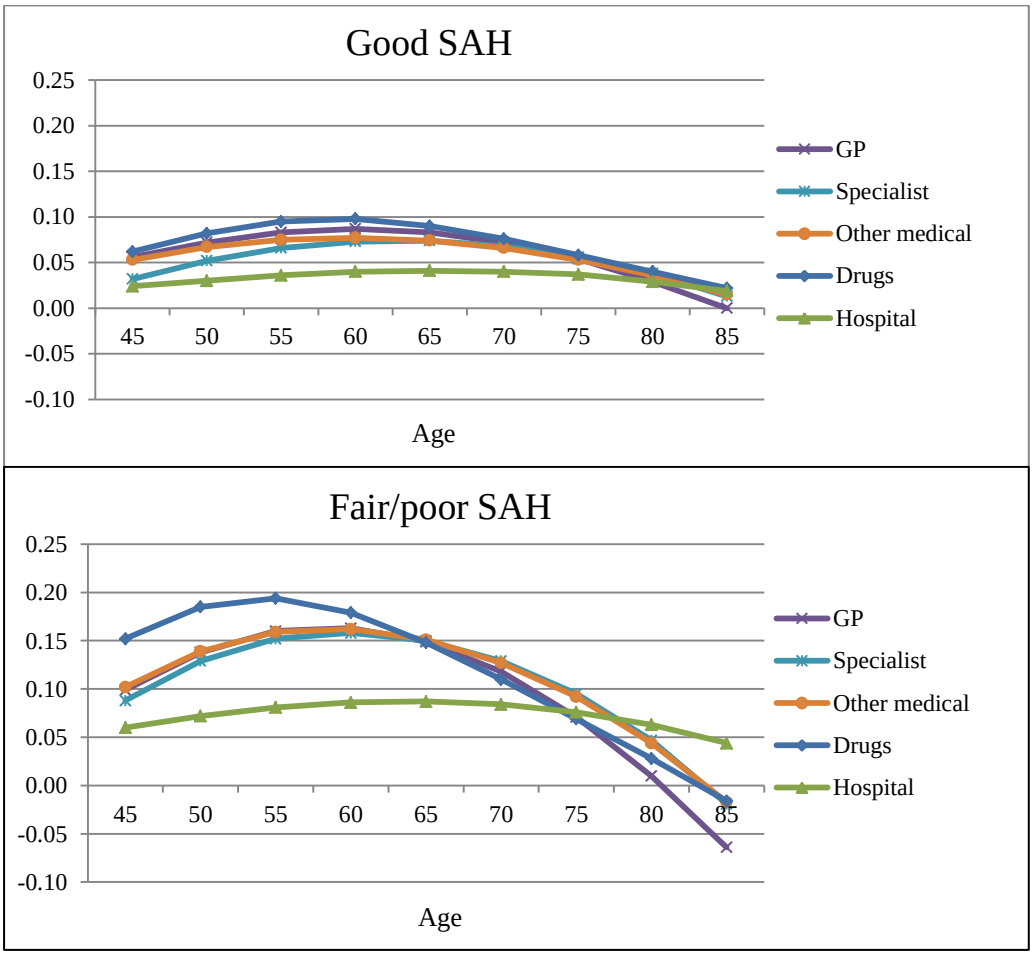
Notes: Sample size is 212,574. Robust standard errors are in parentheses. All logit regressions control for demographic, SES, lifestyle, family health history and QoL variables and year effects. The omitted group for SAH is excellent or very good. Symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 9. Comparison of predictive power of SAH and more objective health measures, pseudo R-squared statistics.

		Dependent Variable				
	Inclusion of health controls	*GP*	*Specialist*	*Other medical services*	*Drugs*	*Hospital*
1	Only non-health controls	0.140	0.057	0.099	0.349	0.054
	Health variables added:					
2	SAH only	0.147	0.064	0.107	0.363	0.058
3	Objective health (survey) only	0.161	0.080	0.124	0.390	0.068
4	Objective health (admin) only	0.246	0.217	0.226	0.636	0.116
5	SAH + Objective health (survey)	0.164	0.082	0.127	0.394	0.069
6	SAH + Objective health (admin)	0.246	0.218	0.227	0.636	0.117
7	Objective health (survey) + Objective health (admin)	0.249	0.220	0.229	0.637	0.118
8	SAH + Objective health (survey) + Objective health (admin)	0.249	0.220	0.230	0.637	0.119

Notes: Sample size is 212,574. All logit regressions control for demographic, SES, lifestyle, family health history, QoL variables and year effects.

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



Notes: Sample size is 212,574. All logit regressions control for demographic, SES, lifestyle, family health history and QoL variables and year effects. The omitted group for SAH is excellent or very good.