

Do I know more than my body can tell?

Information on future health in a self-rated health measure

Amelie C. Wuppermann*

Preliminary Version

March 31, 2009

Abstract

In theory, adverse selection in competitive insurance markets leads to inefficient outcomes. However, empirical evidence on adverse selection is mixed. A reason for this might be that insurance companies can collect enough information about their potential clients so that no important asymmetries in information persist. In this paper, we evaluate whether scope for adverse selection remains in insurances related to health, life and disability when insurance companies have socio-demographic variables, and self-reported and objectively measured health data available for risk assessment. Using a measure of general self-rated health as indicator for private knowledge of health in the English Longitudinal Study of Ageing, we find evidence for significant private knowledge about future health risks and thus scope for adverse selection in the markets for life and health insurance and long-term care.

JEL-Code: D82, I10, C30

Keywords: ELSA, self-rated health, adverse selection, objective health measures

*University of Munich, Kaulbachstr. 45, 80539 Munich, Germany. Email: amelie.wuppermann@lrz.uni-muenchen.de

1 Introduction

Since the seminal papers by Rothschild and Stiglitz (1976) and Wilson (1977) information asymmetries in competitive insurance markets are known to induce inefficiencies due to adverse selection. Extensions of the standard models robustly predict that people with higher risks buy more insurance (Chiappori et al. (2006)). Empirical evidence on this matter is mixed, however. On the one hand, instead of a positive correlation between risk and insurance the opposite is found for many insurance markets in reality (examples are Doiron et al. (2008) and Buchmueller et al. (2008) for private health insurance in Australia, or Fang et al. (2008) for the US Medigap market). On the other hand, UK private health insurance (Olivella et al. (2006)) and UK annuity markets (Finkelstein and Poterba (2002), (2004) and (2006)) are found to suffer from adverse selection.

The mixed empirical evidence on adverse selection is frequently explained by the existence of additional asymmetric information about preferences for insurance (cf. Finkelstein and McGarry (2006), Chiappori et al. (2006) and Cutler et al. (2008)). If people with high preferences for insurance buy insurance and are of lower risk, e.g. because they engage in risk-lowering activities, a negative correlation between risk and insurance coverage arises. This correlation may outweigh the positive correlation resulting from adverse selection.

The indecisive evidence could, however, also result from the fact that insurance companies in some markets employ efficient methods to attenuate the effects of adverse selection. The scope for “active” adverse selection, i.e. selection that results from the fact that buyers of insurance hold private information about their risk and actively use this information to select their insurance coverage (Finkelstein and Poterba (2002)), could be small if the insurance companies manage to extract enough information about their potential clients’ risks so that the clients do not have additional private knowledge. In the case of insurances providing cover for health, life and disability risks for example, the insurance companies typically collect medical information by asking customers to report their prior medical history, prior medical history of their family or even ask customers to take a health screening (for use of medical information in insurances in the UK see ABI and BMI (2008)).

This paper focuses on the question whether scope for adverse selection remains in the markets for health, life and critical illness insurance or long-term care when insurance companies are allowed to collect and use a great amount of variables related to individuals' health risks. In this case, adverse selection is assumed to be possible if individuals know more about their future health risks than the insurance companies can infer from analyzing the health and sociodemographic measures available to them.

As a proxy for private knowledge about future health risks, we use self-rated health (SRH). People were asked to rate their health in general on a five-point-scale from very good to very bad. While this does not directly capture the individual's perception about risk of dying or risk of experiencing a certain disease onset, the measure has been shown to be related to later health events¹.

Differently from many earlier studies, we do not analyze the information that is additional to self-reported and thus subjective health data but also include blood pressure measurement, blood sample analytes, Body Mass Index (BMI) and waist/hip-ratio in the analysis. All these measures are reported by a health official, and are thus considered "objective" health data.²

Using data from the English Longitudinal Study of Ageing (ELSA), we find scope for adverse selection in the case of life insurance and long-term care, as SRH at baseline significantly predicts death and later functional limitations for men and women. We find no evidence, however, that SRH contains information about the onset of specific diseases, like stroke, heart attack or diabetes.

The remainder of the paper is structured as follows. Section II gives an overview on the literature about the predictive power of self-rated health, Section III introduces the data used for the analysis. Section IV outlines the estimation strategy, results are shown in Section V and the last section concludes.

¹For overviews see DeSalvo et al. (2006), Idler and Benyamini (1997), Benyamini and Idler (1999) and the next section.

²The inclusion of objective health data also makes the study interesting from a medical perspective: if SRH contains information about future health risks that is additional to what a physician can measure, it could be an easy to collect and helpful measure for risk assessment (DeSalvo et al. (2006)).

2 Literature on the predictive power of SRH

The predictive power of SRH for all-cause mortality has been analyzed to a great extent in the literature. Idler and Benyamini (1997) review 27 community studies from different countries that analyze the information contained in SRH for future mortality. The authors update their review with 19 additional studies in 1999 (Benyamini and Idler (1999)). All but six studies find that people with lower SRH have higher risks of dying in the next 2 to 13 years after controlling for some other health measures and sociodemographic variables at baseline. DeSalvo et al. (2006) include 20 studies in a meta-analysis, 12 of these had not been reviewed in either of the two papers by Idler and Benyamini. The meta-analysis confirms the significant relationship of SRH and risk of dying even after controlling for functional status, depression and co-morbidities at baseline. People who report poor SRH have an about 2 times higher risk of dying than people who report excellent health.

In addition, more recent studies for the US (Jylhä et al. (2006)), Denmark (Nielsen et al. (2008)) and Canada (Banks et al. (2007)) show similar results. In the studies of Nielsen et al. (2008) and Jylhä et al. (2006) a plethora of objectively measured health data is used as controls. The latter are found to attenuate the relationship between SRH and all-cause mortality but do not render it insignificant.

The relationship between SRH and other future health events has received relatively less attention in the literature. Different studies analyze the relationship between SRH and cause-specific mortality, later disease incidence, functional limitations, cognitive impairment and health service use.

All studies on cause-specific mortality find some evidence for additional predictive power of the self-reported general health measure. The findings are contradictory, however. While Benjamins et al. (2004) find a significant relationship between SRH at baseline and all internal causes of death, that is all causes of death except for accidents and homicide, Pijls et al. (1993) find that death due to cancer and due to all other causes but cancer and cardiovascular disease is predicted by SRH, whereas death due to cardiovascular disease is not. Conversely, Myint et al. (2006) find a significant relationship of self-reported physical

functioning and death due to cardiovascular disease but not due to cancer. All three studies employ cox proportional hazard models and have similar length of follow-up periods (between 5 and 7 years). They vary with respect to included confounding factors at baseline, country of origin and the question that measures self-rated health at baseline.

In the case of disease incidence, the results in the different studies point into similar directions. Pijls et al. (1993) find for the Netherlands that incidences of cardiovascular disease, or cancer are not significantly predicted by SRH. Future diabetes and/or lung disease is marginally significantly related to SRH at baseline. Banks et al. (2007) analyze a Canadian sample. The authors find that having a major disease (cancer, stroke, and/or heart disease) ten years after the baseline interview is significantly related to SRH for females but not for males when controlling for pre-existing conditions, socio-economic variables and some risk factors (like smoking behavior and BMI). Having a medium condition (diabetes and/or hypertension) ten years later is significantly related to SRH for both genders but more strongly so for men. Interestingly, Banks et al. (2007) do not include objectively measured health data while Pijls et al. (1993) use data from an extensive medical examination including anthropometric measures, blood pressure measurement and blood sample analytes. Also, family medical history is included as controls in Pijls et al. (1993). The fact that Banks et al. (2007) find a highly significant relationship between SRH and medium conditions for men while this relationship is only borderline significant in Pijls et al. (1993) might be due to the inclusion of these objective health data. ³

The results concerning the predictive power of SRH for functional limitations or disability of Banks et al. (2007) are in line with what other studies find for different samples using different methods and control variables. SRH at baseline significantly predicts long term activity limitation for men and women in Canada (Banks et al. (2007)). Similar results are found by Idler et al. (2000) and Atchley and Scala (1998) for US-samples and Bond et al. (2006) for the UK. Hillen et al. (2003) find no significant relationship of SRH

³In addition, Banks et al. (2007) do not use a Cox proportional hazard model but logistic regression and analyze a younger sample. The use of different models, however does not lead to differing results when analyzing the predictive power of SRH for all-cause mortality (Benyamini and Idler (1999)) or in the case of functional limitations.

and later functional limitations using a different UK-sample. While the other studies are all community studies, Hillen et al. (2003) analyze stroke patients.

The similarity of the results in the community studies is especially interesting as all studies use different definitions of functional limitations.⁴

Later cognitive impairment is not significantly predicted by SRH when controlling for other health measures at baseline (Bond et al. (2006)). Similarly, later use of health services is only found to be significantly predicted when not controlling for health measures at baseline (Miilunpalo et al. (1997) and Dening et al. (1998)). Bath (1999) includes baseline service use and medication which render the predictive power of SRH for later service use insignificant.

Overall, the existing literature shows that SRH contains significant additional information to predict later health events. While most of the studies analyzing the predictive power for all-cause mortality consistently find a significant relationship, other health events have been analyzed to a lesser extent and the results are not as clear cut. It is thus interesting to shed more light on the predictive power of SRH for different later health events. In addition, almost all studies report that they suffer from attrition and missing data but missings are just ignored in most cases.⁵ In the following, we try to evaluate the possibility of bias due to item non-response and attrition and employ inverse probability weighting as a correction where bias seems likely.

⁴Banks et al. (2007) classify individuals as “long term restricted” if they are limited in the kind or amount of activity they can perform at home, school, work or other because of a physical or mental problem. Idler et al. (2000) use a summary score of 23 items measuring function in 8 domains from the Stanford Health Assessment Questionnaire. Atchley and Scala (1998) create functional limitation scores that count for how many of the three items “walking half a mile”, “walking up and down stairs” and “doing heavy work around the house” individuals say they are unable to perform. Bond et al. (2006) classify people as functionally impaired if they show an inability to do activities of daily living (ADL), Hillen et al. (2003) count people as limited in their functional ability if they have a Barthel index below 20.

⁵A notable exception are Banks et al. (2007) who conduct a variable addition test to evaluate how attrition affects their estimates.

3 Data

The English Longitudinal Study of Ageing (ELSA) is a rich panel data set which contains socio-demographic, economic and health related data on at most 11,392 individuals that were born on or before February 29th, 1952 and were living in private homes in England at the time of the interview. In addition to those core sample members younger partners living in the same household are interviewed as part of ELSA. The sample was randomly selected from the English population in three repeated cross sections for the Health Survey for England (HSE) in the years 1998, 1999 and 2001. In addition to the data from the eligible ELSA subsample of the three HSE years, called ELSA wave 0, data is available from three ELSA waves collected in 2002, 2004 and 2006.

While the design of ELSA was highly influenced by and modelled on the US Health and Retirement Study (HRS), it varies in one important feature: In addition to the biannual interview data, every four years a nurse visit is conducted as part of ELSA. Due to this nurse visit objectively measured blood pressure, blood sample analytes and anthropometric data are available. Up to now data on nurse visits is available for wave 0 and wave 2. In wave 0, however, a blood sample analysis was only undertaken for all consenting sample members from the HSE year 1998. As it is especially interesting to see whether private information about future health risks remains even when insurance companies are allowed to use blood analytes, we concentrate on ELSA sample members that were sampled for the 1998 HSE.

The ELSA data in wave 0 contains 8,267 people from HSE 1998 (7,807 core sample members and 459 younger partners). As everyone who conducted the interview was eligible for the nurse visit in HSE, core sample members and younger partners from wave 0 are included in this analysis. While everyone was eligible for the nurse visit, not everyone actually had a nurse visit conducted as people could refuse the latter.

Overall, only for 3,956 individuals from HSE 1998 all subjective and objective health measures are available. Subjective health data on co-morbidities is missing for 23 observations. These observations are deleted. For another 4,288 observations either valid blood

pressure measurement, valid blood sample analytes or valid BMI or waist/hip-ratio are not available. As can be seen in columns 1 to 3 of Tables 2 and 3 the differences in means between people with and people without objective health data are significant for most of the variables. The people, for whom objective health data is missing, seem to be older, less well educated, and poorer. More of them smoke and less are married. The sample without objective health measures also reports lower SRH and is less healthy with respect to self-reported conditions.

Of course, this raises the worry, that estimating the predictive power of SRH just on the sample for which objective health measures are available might produce biased results. Several approaches will be taken to evaluate how important the worry is in this data.

The sample size shrinks further due to attrition when analyzing the incidence or onset of later specific conditions other than death. Information on death is collected regardless of attrition for everyone who gave consent by linking data to information from the Department of Work and Pensions and to data about death contained in the National Health Service Central Register held by the Office of National Statistics. Information on later special conditions, however, is only available, if people appear again in ELSA after wave 0. As some people are not observed in one wave but occur again in later waves, we analyze the onset of diseases or functional limitation any time after wave 0 instead of looking at one specific later wave in order to lose as few observations as possible.^{6, 7} Still, of the 3,956 people with objective health data only 71 percent are observed at least once after wave 0. The others drop out because they refuse further participation (56.9 percent of attrition), because they die (15.5 percent) or for other reasons.

As can be seen in columns 4 to 6 of Tables 2 and 3 and in Table 4 the people who are not observed again after wave 0 vary in many aspects from the others. They are on average older, less well educated, poorer, less likely to be white or married and more likely to be smokers. Also, they have worse outcomes in both subjective and objective health

⁶For the coding of diseases see Table 1.

⁷For the 3,956 people with objective health data, 16 people are not observed in wave 1 but observed in waves 2 and/or 3.

measures.

The estimates of the predictive power of SRH for later health events might therefore not only be biased by missing objective health data but also by attrition. As a first approach, the results are nevertheless presented for the selected samples. The possibility of biases will be evaluated by different methods.

Figures 1 and 2 present a first glance into the relationship between SRH and later health events. There seems to be a graded relationship of SRH and all-cause mortality for both genders: The better SRH in 1998, the lower the average of people who are known to be dead by the end of 2005. For men the same is true for cardiovascular conditions, diabetes, the incidence of functional limitations, longstanding illness, and major conditions. The relationship of SRH with cancer, arthritis or being unable to work due to illness is less distinct. For women, a graded relationship of SRH can be seen with angina, heart attack, functional limitations and longstanding illness. The other health events are not systematically related to SRH at baseline.

To evaluate whether the observed relationships also hold when controlling for co-morbidities and other health measures at baseline, a more thorough analysis is conducted. The estimation strategy is outlined in the next section.

4 Estimation

As a first approach, we estimate the relationship between SRH and later health events using single equation probit estimation on the selected samples for men and women separately. Let y_i^{j*} , $j \in \{1, 2, \dots, J\}$ denote latent variables for each of the J future health events for individual i . Each of the y_i^{j*} 's can be captured by the following equations

$$\begin{aligned} y_{i,t+k}^{j*} &= \beta_0^j + \beta_1^j SRH_{it} + \beta_2^j OH_{it} + \beta_3^j SH_{it} + \beta_4^j c_{it} + \epsilon_{it}^j \\ y_{i,t+k}^j &= \mathbb{I}(y_{i,t+k}^{j*} > 0) \end{aligned} \quad (1)$$

where OH_{it} is a vector that includes the objective health variables, the vector SH_{it} captures all self-reported health measures and the vector c_{it} stands for the socio-demographic

variables age, education, income, marital status and employment status. ϵ_{it}^j captures unobservables influences on the latent later health event j .

Under the assumptions that $\epsilon_{it}^j \sim N(0, 1)$ and that the correlation between ϵ_t^j and ϵ_t^k , ρ_{jk} , is equal to 0 for all $j \neq k$, we can estimate $Pr(y_{i,t+k}^j = 1)$ for each of the J health events independently using a single equation probit model.⁸

All estimations are conducted on the selected samples. In general, only people are included for whom all objective health data are observed. When analyzing the onset of a specific disease as dependent variable only people who are observed in ELSA at least once after wave 0 — and for whom the specific question is not missing — are included.

Whether the estimated coefficients are consistent estimates for the population depends crucially on the selection mechanisms that determine the availability of all health measures and attrition. Consistent estimates are produced under the assumption that the distribution of the later health event given the covariates and a selection indicator is the same as the distribution of the later health events conditional only on the covariates (Imbens and Wooldridge (2008), p. 807).

To see whether the selection mechanism that results in missing objective health data can potentially bias our estimates, a variable addition test is conducted. Instead of the objective health measures an indicator for selection is included in equation (1) and the equation is estimated on the whole sample. A significant estimate for the coefficient of the selection indicator hints at differences in the relationship between SRH and the later health event for the two samples conditional on all other covariates (except the objective health data).

Whether attrition biases the results is tested by another variable addition test a la Verbeek and Nijman (1992). Later health events are defined only with data from wave 1. An indicator capturing whether people are still observed in wave 2 is additionally included

⁸The assumption of independence between the error terms could be relaxed, by estimating multiple equation probit models. In this case, the correlation between the different error terms is unrestricted and is estimated jointly with the parameters of the model by maximum likelihood using the Geweke-Hajivassiliou-Kaene (GHK) simulator to evaluate the loglikelihood (see Cappellari and Jenkins (2003)). The benefit of estimating the equations jointly would merely be an efficiency gain. It is not pursued further.

in equation (1). Again, a significant coefficient for the indicator hints at differences in the relationship between SRH and the outcome of interest for people who were observed in wave 0 and 1 but drop out after wave 2 and people who are still observed in wave 2.

One possibility to correct potentially arising biases due to missings and attrition is inverse probability weighting. Intuitively, the probability of being in the sample is estimated and the selected sample is then weighted with the inverse of the estimated probability.⁹ This way, observations that are observed despite a low probability of being in the sample get a high weight, as they are under-represented in the selected sample.

We employ two different weightings to correct the two selection mechanisms. For the missing objective health data, a strong form of the selection on observables assumption has to hold in order for the weighted estimates to help correct the bias. We need to assume, that $Pr(s = 1|y, z, OH) = Pr(s = 1|z)$, where s is a selection indicator, and z captures all control variables in equation (1) except for objective health data. In addition z includes at least one variable that is related to the outcome of interest y but is deliberately excluded from the equation of interest (Jones et al. (2007), p. 274).

In the case of attrition, we need the assumption $Pr(a = 1|y, x) = Pr(a = 1|y)$, where a is an indicator that takes on the value 1 for people that are observed again after wave 0, and 0 for people who drop out after wave 0. x now captures all variables in equation (1), plus again at least one variable that is related to the outcome of interest, y , but excluded from equation (1).

We thus need two exclusion restrictions for the two different weighting procedures. For missing objective health data this is an indicator that captures if people were interviewed on their own or together with other people. This is assumed to predict the availability of objective health data, as social pressure could make people more likely to consent to a nurse visit when other people are around during the interview. In addition, being interviewed alone is likely correlated with the later health outcomes, as health at baseline could affect the number of people who are present during the interview. It should, however, not have

⁹The fact that the weight is ignored in our analysis. This leads to conservative inference (Imbens and Wooldridge (2008), p.817).

an effect on the later health outcomes that is in addition to the included health measures and it is thus excluded from equation (1).

For attrition, we use whether people rent or do not rent their own apartment/house at the time of the baseline interview as exclusion. While this could be a proxy for wealth at baseline, and thus be related to health aswell, we assume that we can exclude it from equation (1) as socio-economic variables (education and income) at baseline are already included. Renting or not is also a measure which is related to attrition, as at least part of the attrition is due to people moving and not being found again (about 7 percent). We conjecture that changing address is likely related to the state of ownership of the house.

With these two different weighting strategies, the weighting used to correct one selection mechanisms relies on the assumption that the other mechanisms is at random and does not bias our estimates. Ideally, the two selection mechanisms should be modelled jointly like in Jenkins and Cappellari (2004). This is on our agenda for the future.

5 Results

The results of the single equation probit models estimated on the selected samples are displayed in Tables 6 and 5 for men and women, respectively. Both tables show marginal effects for SRH being very bad or bad, good or very good compared to fair SRH. For each later health event, the first row includes all baseline controls except for objective health data, the second row also includes blood pressure measurement, BMI, waist/hip-ratio and blood sample analytes.

Looking at the results that include objective health data, it becomes clear that SRH contains additional information for many of the future health events for both genders.

For men, being limited in activities of daily living, an onset of major conditions, a general longstanding illness and being unable to work due to illness sometime during the 8 years of follow-up is significantly predicted by SRH at baseline. Looking at the specific diseases, however, the relationship is only rarely significant for men. In the case of psychiatric problems it even turns out that people in bad or very bad SRH at baseline are less

likely to experience an onset compared to people in fair SRH.

For women, the information contained in SRH seems to be a little different. While a general onset of one of the major conditions is not significantly related to SRH at baseline when controlling for all other baseline controls, an onset of some of the specific conditions (angina and arthritis) is significantly predicted by SRH. Similar to men, functional limitation, longstanding illness and being unable to work due to illness are predicted by SRH.

In case of death, the marginal effects show the expected signs for all SRH states for both men and women. For men, however, only being in very good SRH compared to fair SRH is significantly related to lower probability of dying. For women, all coefficients are significant.

Interestingly the inclusion of objective health data does not greatly vary the coefficients as can be seen from comparisons of the lines a and b in Tables 6 and 5. The tendency seems to be however, that the inclusion of objective health data lowers the point estimates (in absolute value) and reduces their significance levels. This suggests that the availability of objective health data to insurances reduces the scope for adverse selection.

The results interpreted so far are all achieved analyzing the selected samples. Table 7 presents results on the variable addition tests conducted to evaluate the worry, that the two selection mechanisms might bias the results. For most of the later health events neither of the two selection mechanisms seems to be a great challenge.

The selection mechanism that leads to missing data on objective health is evaluated in the first column. In probit estimations on the whole sample explaining later health events by all control variables except for objective health data and an indicator for the availability of objective health data, this indicator is hardly significant. For the cases where it is significant, the inclusion of objective health data, were they available for everyone, could eliminate or at least reduce the significant differences between the two samples. This is especially true, as the sample with objective health data is more healthy in terms of self-reported data than the sample without objective health data and might thus also be

more healthy in terms of objective health measures. It is therefore likely that the selection mechanism which results in missing objective health data does not pose a great worry. To further underline this finding, we correct the possibly arising bias using inverse probability weighting.

In the last two columns of Table 7, the possibility that attrition affects the results is evaluated. Most of the different health outcomes do not vary significantly between people still observed in wave 2 and people who drop out after wave 1. However, for onset of cancer, the incidence of stroke and being unable to work due to illness, the indicator is significant for women. Therefore inverse probability weighting is employed for attrition as well.

The weighted estimates reported in Tables 9 and 8 are very similar to the unweighted estimates. Interestingly, the estimates with weights are overall more often significant and the marginal effects are higher in absolute values than the estimates without weights. This suggests, that the relationship between SRH and later health events is stronger for the people for whom we do not observe objective health data or for those people who drop out, and thus that the estimates on the selected sample seem to be conservative. As the selected sample is on average healthier than the overall sample, SRH might be a more informative measure for sick people.

It has to be kept in mind, however, that the weighting to correct for one selection mechanism assumes that the other selection mechanism can be ignored. This assumption will be relaxed by jointly modelling the two selection mechanisms.

Overall, the results suggest that there is scope for active adverse selection in the UK market for long term care: The incidence of later functional limitation, that is being limited in the activities of daily living, is robustly predicted for men and women by SRH. Adverse selection seems also possible in the market for life insurance. This is in line with the results of Finkelstein and Poterba (2002), (2004) and (2006) who find evidence for adverse selection in the UK annuity market. We additionally find evidence that the market of general private health insurance might potentially suffer from adverse selection, as SRH contains information on having longstanding illnesses in the future. This result nicely

aligns with Olivella and Vera-Hernandez (2006). We do not find evidence for potential adverse selection in the market of critical illness coverage. The onset of specific conditions is hardly predicted by SRH. It has to be born in mind, however, that SRH is only a proxy for private knowledge about future health risks. Individual assessments of the risk of a certain disease onset could be more helpful for this matter, especially as there could be knowledge about genetic disposal to certain diseases which need not necessarily be reflected in SRH.

6 Conclusion

In this paper, we evaluate the scope for adverse selection in health related insurance markets by analyzing how well self-rated health predicts future health events when information on self-reported and objective health measures is taken into account.

Analyzing data from the English Longitudinal Study of Ageing we find evidence for the possibility of adverse selection in the UK markets for long-term care and life insurance. Even if insurance companies can use objectively measured blood pressure, anthropometric measures and outcomes of a blood sample analysis in addition to self-reported conditions, family history and some socio-economic and demographic variables like age, gender, education and income, there are systematic factors captured by SRH that contain additional information for death and disability. The same is true for general health insurances, as SRH contains additional information for having a longstanding illness.

In contrast, SRH contains no additional information for the onset of specific conditions, however. This casts doubt on the scope for adverse selection in critical illness insurance. SRH might, however, not be the best available proxy to measure private knowledge, as for example genetic disposition to a certain disease might be known to an individual but this knowledge need not influence her self-rated health state.

While we find that there is scope for adverse selection in some markets, we can show that this scope is likely reduced by the availability of objective health data. Had the insurance company even more data available (for example additional outcomes of the blood sample

analysis), the scope for adverse selection could diminish further. More light could be shed on this issue by analyzing later waves of ELSA for which more objective health data, like measures of grip strength and other blood sample analytes, are available.

Another important aspect of our study is that we evaluate whether selection due to attrition or due to missing objective health data biases our estimates. We come to the conclusion, that selection does not seem to play an important role. However, we plan to shed more light on this issue by explicitly modelling both selection mechanisms jointly.

References

- ABI and BMI (2008). Medical information and insurance. Joint guidelines, Association of British Insurers and British Medical Association, www.abi.org.uk.
- Atchley, R. and Scala, M. (1998). Long-range antecedents of functional capability in later life. *Journal of Aging and Health*, 10(1):3–19.
- Banks, J., Crossley, T., and Goshev, S. (2007). Looking for Private Information in Self-assessed Health. HEDG Working Paper, University of York 9.
- Bath, P. (1999). Self-rated health as a risk factor for prescribed drug use and future health and social service use in older people. *Journals of Gerontology Series A: Biological and Medical Sciences*, 54(11):565–570.
- Benjamins, M., Hummer, R., Eberstein, I., and Nam, C. (2004). Self-reported health and adult mortality risk: an analysis of cause-specific mortality. *Social Science & Medicine*, 59(6):1297–1306.
- Benyamini, Y. and Idler, E. (1999). Community studies reporting association between self-rated health and mortality: additional studies, 1995 to 1998. *Research on Aging*, 21(3):392–401.
- Bond, J., Dickinson, H., Matthews, F., Jagger, C., and Brayne, C. (2006). Self-rated health status as a predictor of death, functional and cognitive impairment: a longitudinal cohort study. *European Journal of Ageing*, 3(4):193–206.
- Buchmueller, T., Fiebig, D., J. G., and Savage, E. (2008). Advantageous Selection in Private Health Insurance: The Case of Australia. CHERE Working Paper 2008-2.
- Cappellari, L. and Jenkins, S. P. (2003). Multivariate probit regression using simulated maximum likelihood. *Stata Journal*, 3(3):278–294.

- Chiappori, P., Jullien, B., Salanie, B., and Salanie, F. (2006). Asymmetric information in insurance: General testable implications. *Rand Journal of Economics*, 37(4):783–798.
- Cutler, D., Finkelstein, A., and McGarry, K. (2008). Preference heterogeneity and insurance markets: Explaining a puzzle of insurance. *American Economic Review*, 98(2):157–162.
- Dening, T., Chi, L., Brayne, C., Huppert, F., Paykel, E., and O’Connor, D. (1998). Changes in self-rated health, disability and contact with services in a very elderly cohort: a 6-year follow-up study. *Age and Ageing*, 27(1):23–33.
- DeSalvo, K., Bloser, N., Reynolds, K., He, J., and Muntner, P. (2006). A Meta-Analysis. *Journal of General Internal Medicine*, 21(3):267–275.
- Doiron, D., Jones, G., and Savage, E. (2008). Healthy, wealthy and insured? The role of self-assessed health in the demand for private health insurance. *Health Economics*, 17(3):317–34.
- Fang, H., Keane, M., and Silverman, D. (2008). Sources of advantageous selection: Evidence from the Medigap insurance market. *Journal of Political Economy*, 116(2):303–350.
- Finkelstein, A. and McGarry, K. (2006). Multiple dimensions of private information: evidence from the long-term care insurance market. *The American Economic Review*, 96(4):938–958.
- Finkelstein, A. and Poterba, J. (2002). Selection effects in the United Kingdom individual annuities market. *The Economic Journal*, 112(476):28–50.
- Finkelstein, A. and Poterba, J. (2004). Adverse Selection in Insurance Markets: Policyholder Evidence from the UK Annuity Market. *Journal of Political Economy*, 112(1):183–208.

- Finkelstein, A. and Poterba, J. (2006). Testing for Adverse Selection with Unused Observables. *NBER Working Paper*, (12112).
- Herold, G. (2004). *Innere Medizin*. Gerd Herold, Köln, Germany.
- Hillen, T., Davies, S., Rudd, A., Kieselbach, T., and Wolfe, C. (2003). Self ratings of health predict functional outcome and recurrence free survival after stroke. *British Medical Journal*, 57(12):960–966.
- Idler, E. and Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav*, 38(1):21–37.
- Idler, E., Russell, L., and Davis, D. (2000). Survival, functional limitations, and self-rated health in the NHANES I Epidemiologic Follow-up Study, 1992. *American Journal of Epidemiology*, 152(9):874–883.
- Imbens, G. and Wooldridge, J. (2008). Lecture Note 18: Missing Data. University of wisconsin-madison, institute for research on poverty.
- Jenkins, S. and Cappellari, L. (2004). Modelling low pay transition probabilities, accounting for panel attrition, non-response and initial conditions. CESifo Working Paper 1232, ifo Institute, Munich.
- Jones, A., Rice, N., Bago d’Uva, T., and Balia, S. (2007). *Applied health economics*. Routledge.
- Jylhä, M., Volpato, S., and Guralnik, J. (2006). Self-rated health showed a graded association with frequently used biomarkers in a large population sample. *Journal of Clinical Epidemiology*, 59(5):465–471.
- Miilunpalo, S., Vuori, I., Oja, P., Pasanen, M., and 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.

- Myint, P., Luben, R., Surtees, P., Wainwright, N., Welch, A., Bingham, S., Day, N., Wareham, N., and Khaw, K. (2006). Relation Between Self-Reported Physical Functional Health and Chronic Disease Mortality in Men And Women in the European Prospective Investigation Into Cancer (EPIC–Norfolk): A Prospective Population Study. *Annals of Epidemiology*, 16(6):492–500.
- Nielsen, S., Brit, A., Siersma, V., Hiort, C., et al. (2008). Self-rated general health among 40-year-old Danes and its association with all-cause mortality at 10-, 20-, and 29 years' follow-up. *Scandinavian Journal of Public Health*, 36(1):3–11.
- Oliveira, C. (2008). Nurse Visit - ELSA user day. Presentation at elsa user day 2008, Economic and Social Data Service.
- Olivella, P. and Vera-Hernandez, M. (2006). Testing for adverse selection into private medical insurance. Institute for Fiscal Studies Working Paper 06/02.
- Pijls, L., Feskens, E., and Kromhout, D. (1993). Self-rated Health, Mortality, and Chronic Diseases in Elderly Men The Zutphen Study, 1985-1990. *American Journal of Epidemiology*, 138(10):840–848.
- Rothschild, M. and Stiglitz, J. (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *Quarterly Journal of Economics*, 90(4):629–649.
- Verbeek, M. and Nijman, T. (1992). Testing for selectivity bias in panel data models. *International Economic Review*, 33(3):681–703.
- Wilson, C. (1977). A Model of Insurance Markets with Incomplete Information. *Journal of Economic Theory*, 16(2):167–207.

Tables

Table 1: Coding of health events

Health Event	Description
Death	1 if person known to be dead by end of 2005
	0 if alive or not known to be dead by end of 2005
Cancer, Diabetes Angina, Arthritis Psychiatric Problem	1 if individual reports having been diagnosed with condition at an age older than age at baseline interview and/or at a point in time after baseline interview
	0 otherwise
Heart Attack, Stroke	1 if either first diagnosis after baseline interview (as above) or if additional occurrence reported after baseline interview
	0 otherwise
Major Condition	1 if Cancer, Angina, Heart Attack and/or Stroke
	0 if neither
Functional Limitation	1 if at least one limitation in activity of daily living (bathing or showering, dressing, eating, walking across room, getting in and out of bed, using toilet) reported at least once after baseline interview
	0 otherwise
Longstanding Illness	1 if longstanding illness reported at least once after baseline interview
	0 otherwise
Unable to work	1 if individual reports not being able to work due to illness at least once after baseline interview
	0 otherwise

Table 2: Descriptives - ELSA Wave 0

Variable	A Mean	B Mean	(A - B) Difference	C Mean	D Mean	(C - D) Difference
Age <50	.15	.21	-.062***	.21	.21	-.001
Age 50-59 ^a	.27	.32	-.049***	.29	.34	-.051***
Age 60-69	.25	.25	-.002	.20	.27	-.067***
Age 70-79	.22	.16	.064***	.18	.14	.037***
Age 80-89	.09	.06	.035***	.11	.04	.074***
Age > 90	.02	.004	.013***	.01	.00	.007***
Male	.42	.45	-.029***	.46	.44	.023
University degree	.08	.10	-.021***	.08	.11	-.027**
University without degree	.08	.11	-.023***	.09	.11	-.030***
A-level	.05	.06	-.012**	.06	.060	.003
O-level	.13	.17	-.044***	.15	.18	-.034**
NVQ1 equivalent	.04	.05	-.004	.06	.04	.022***
Other qualification	.08	.07	.009	.064	.07	-.009
No qualification ^a	.53	.44	.095***	.49	.42	.075***
Student	.02	.02	-.000	.021	.02	-.002
1st income quintile ^a	.23	.18	.054***	.23	.16	.065***
2nd income quintile	.14	.13	.013*	.13	.13	.003
3rd income quintile	.17	.20	-.028***	.18	.20	-.021
4th income quintile	.12	.16	-.045***	.14	.18	-.037***
5th income quintile	.14	.20	-.058***	.17	.21	-.040***
White	.96	.97	-.011***	.96	.98	-.022***
Married	.64	.72	-.080***	.68	.74	-.064***
Unemployed	.02	.02	.001	.02	.02	-.001
Current smoker	.24	.19	.050***	.22	.18	.041***
Ever smoked	.65	.63	.019*	.66	.61	.049***
Number of observations	3,956	4,288		1,141	2,815	

* p<0.10, ** p<0.05, *** p<0.01

Notes:

A - Sample without objective health data,

B - sample with objective health data

C - part of B not observed again after wave 0

D - part of B observed at least once after wave 0

^a Reference category in analysis.

All variables are binary, t-test for difference in means

Table 3: Descriptives - Self-reported Health Measures Wave 0

Variable	A Mean	B Mean	(A - B) Difference	C Mean	D Mean	(C - D) Difference
Self Rated Health						
SRH bad/very bad	.13	.07	.084***	.09	.06	.026***
SRH fair	.28	.21	.068***	.23	.20	.031**
SRH good	.35	.41	-.055***	.40	.41	-.009
SRH very good	.24	.32	-.081***	.28	.33	-.0480***
Diagnosed diseases						
Hypertension	.35	.24	.116***	.25	.23	.020
Diabetes	.04	.03	.017***	.05	.02	.028***
Stroke	.04	.01	.0313***	.02	.01	.016***
Heart attack	.05	.02	.029***	.04	.009	.036***
Angina	.09	.03	.054***	.06	.02	.043***
Heart murmur	.04	.03	.011***	.04	.03	.005
Irr. heart rhythm	.09	.05	.035***	.06	.05	.011
Oth. heart problem	.03	.01	.024***	.01	.01	.002
Other self reported health measures						
Longstanding illness ¹	.61	.51	.100***	.53	.50	.024
Funct. limit	.42	.30	.120***	.33	.28	.043***
Unable to work	.07	.04	.030***	.04	.04	-.002
stroke cause of mother's death	.09	.08	.008	.08	.08	.001
stroke cause of father's death	.07	.07	.006	.07	.07	-.000
heart attack cause of mother's death	.13	.10	.027***	.10	.10	-.003
heart attack cause of father's death	.18	.19	-.008	.17	.19	-.019
Number of observations	3,956	4,288		1,141	2,815	

* p<0.10, ** p<0.05, *** p<0.01

Notes:

A - Sample without objective health data,

B - sample with objective health data

C - part of B not observed again after wave 0

D - part of B observed at least once after wave 0

All variables are binary, t-test for differences in means

¹ Longstanding illness includes: Cancer, Diabetes, Other endocrine/metabolic, Mental illness, Mental handicap, Epilepsy/fits/convulsions, Migraine/headaches, Other problems of nervous system, Cataract/poor eye sight/blindness, Other eye complaints, Poor hearing/deafness, Tinnitus, Meniere's disease/ear complaints causing balance problems, Other ear complaints, Stroke/cerebral haemorrhage/cerebral thrombosis, Heart attack/angina, Hypertension/high blood pressure, Other heart problems, Piles/haemorrhoids, Other blood vessels/embolic, Bronchitis/emphysema, Asthma, Hayfever, Other respiratory complaints, Stomach ulcer/abdominal hernia/rupture, Other digestive complaints, Complaints of bowel/colon, Complaints of teeth/mouth/tongue, Kidney complaints, Urinary tract infection, Other bladder problems/incontinence, Reproductive system disorders, Arthritis/rheumatism/fibrositis, Back problems/slipped disc/spine/neck, Other problems of bones/joints/muscles, Infectious and parasitic disease, Disorders of blood and blood forming organs and Skin complaints.

Table 4: Objective Health Measures in ELSA wave 0

Variable	Overall	Attrition	No attrition	(C - D) Difference
	B Mean	C Mean	D Mean	
Obesity				
BMI \geq 30	.22	.22	.23	-.007
25<BMI<30	.45	.43	.46	-.026
Waist-hip/ratio $>1^a$, 0.85 ^b	.20	.20	.20	.004
Blood pressure				
Systolic BP >140	.46	.51	.45	.061***
Diastolic BP >85	.26	.29	.24	.046***
Both levels evaluated	.22	.25	.20	.051***
Blood sample analysis¹				
Haemoglobin $<13^a$ 11.5 ^b g/dL	.06	.08	.05	.030***
Haemoglobin $>18^a$ 16.5 ^b g/dL	.002	.004	.001	.004**
Ferritin $>400^a$, 200 ^b μ g/L	.03	.04	.03	.010*
Ferritin $<25^a$, 20 ^b μ g/L	.10	.10	.10	.000
Total cholesterol >5 mmol/L	.81	.79	.81	-.024*
HDL cholesterol $<1^a$, 1.2 ^b mmol/L	.22	.23	.21	.021
C-reactive protein >5 mg/L	.20	.23	.19	.039***
Fibrinogen <1.7 g/L	.02	.01	.02	-.004
Fibrinogen >3.7 g/L	.10	.12	.09	.025**
Number of observations	3,956	1,141	2,815	

* p<0.10, ** p<0.05, *** p<0.01

Notes:

B - sample with objective health data

C - part of B not observed again after wave 0

D - part of B observed at least once after wave 0

¹ Reference ranges taken from Oliveira (2008), for people older than 50, where ranges differ by age, values for younger groups taken from Herold (2004)^a Value for males^b Value for females^c Reference ranges vary also by age group, shown ranges for people 46+

Table 5: Probit marginal effects of SRH for later health events: Women

Health Event		SRH – very bad/bad		SRH – good		SRH – very good	
Dead (all causes)	a	0.041**	(0.024)	-0.023**	(0.009)	-0.026**	(0.009)
	b	0.037**	(0.023)	-0.020**	(0.009)	-0.022**	(0.009)
Functional Limitation	a	0.185***	(0.072)	-0.130***	(0.030)	-0.187***	(0.030)
	b	0.182***	(0.072)	-0.135***	(0.030)	-0.179***	(0.031)
Major Condition	a	-0.023	(0.032)	-0.022	(0.021)	-0.018	(0.023)
	b	-0.029	(0.030)	-0.021	(0.020)	-0.013	(0.023)
Cancer	a	-0.031*	(0.008)	-0.016	(0.013)	0.009	(0.015)
	b	-0.029***	(0.006)	-0.014	(0.012)	0.011	(0.014)
Angina	a	0.010	(0.020)	-0.015*	(0.008)	-0.024***	(0.008)
	b	0.010	(0.018)	-0.014**	(0.007)	-0.021***	(0.007)
Stroke	a	-0.001	(0.007)	0.003	(0.004)	-0.002	(0.004)
	b	-0.001	(0.002)	0.001	(0.001)	-0.001	(0.002)
Diabetes	a	-0.002	(0.013)	-0.007	(0.007)	-0.014*	(0.007)
	b	-0.002	(0.001)	-0.002	(0.002)	-0.001	(0.002)
Arthritis	a	0.037	(0.062)	-0.038	(0.028)	-0.057*	(0.029)
	b	0.025	(0.059)	-0.037	(0.028)	-0.054*	(0.029)
Psychiatric Problem	a	0.011	(0.029)	-0.033***	(0.012)	-0.051***	(0.012)
	b	0.009	(0.025)	-0.029***	(0.011)	-0.046***	(0.011)
Longstanding Illness	a	0.207*	(0.074)	-0.093**	(0.044)	-0.229***	(0.049)
	b	0.211*	(0.072)	-0.089**	(0.044)	-0.223***	(0.049)
Unable to work	a	0.043***	(0.023)	-0.018***	(0.007)	-0.011*	(0.006)
	b	0.035***	(0.019)	-0.013***	(0.006)	-0.007	(0.005)

* p<0.10, ** p<0.05, *** p<0.01 - test of the underlying coefficient being 0

Notes: a: without objective health data, b: with objective health data. Reference category: SRH – fair. Marginal effects calculated as $\Phi(\bar{x}\hat{\beta} + \hat{\beta}_{SRH}) - \Phi(\bar{x}\hat{\beta})$ at the mean of all other variables, \bar{x} . Robust standard errors in parentheses. Objective health data: Blood analytes (haemoglobin, ferritin, total cholesterol, hdl cholesterol, c-reactive protein, fibrinogen), systolic and diastolic blood pressure, bmi, waist/hip-ratio. Included controls in all estimations: Age groups, income quintiles, dummies for education outcomes, race, employment status, marital status, self-reported conditions at baseline, functional limitations at baseline, unable to work at baseline, smoking status, parents cause of death (stroke or heart attack). Heart Attack not reported as less than 2 percent of women experienced this health event.

Table 6: Probit marginal effects of SRH for later health events: Men

Health Event		SRH – very bad/bad		SRH – good		SRH – very good	
Dead (all causes)	a	0.044	(0.035)	-0.018	(0.018)	-0.040**	(0.019)
	b	0.032	(0.034)	-0.018	(0.018)	-0.039**	(0.018)
Functional Limitation	a	0.140**	(0.077)	-0.130***	(0.034)	-0.180***	(0.035)
	b	0.131*	(0.083)	-0.116***	(0.036)	-0.168***	(0.037)
Major Condition	a	0.056	(0.052)	-0.074***	(0.027)	-0.077***	(0.028)
	b	0.051	(0.056)	-0.070**	(0.028)	-0.078**	(0.029)
Cancer	a	0.010	(0.033)	0.003	(0.018)	0.006	(0.019)
	b	0.010	(0.033)	0.001	(0.015)	0.003	(0.017)
Heart Attack	a	-0.003	(0.020)	-0.010	(0.012)	-0.019	(0.013)
	b	0.003	(0.022)	-0.005	(0.012)	-0.013	(0.013)
Angina	a	-0.011	(0.015)	-0.024*	(0.013)	-0.024*	(0.013)
	b	-0.003	(0.015)	-0.017	(0.011)	-0.015	(0.012)
Stroke	a	0.005	(0.009)	0.002	(0.005)	-0.002	(0.005)
	b	0.008	(0.011)	0.003	(0.004)	0.001	(0.005)
Diabetes	a	0.028	(0.043)	-0.008	(0.020)	-0.026	(0.022)
	b	0.047	(0.046)	-0.001	(0.016)	-0.006	(0.018)
Arthritis	a	-0.054	(0.029)	-0.024	(0.027)	-0.043	(0.029)
	b	-0.053	(0.029)	-0.018	(0.027)	-0.039	(0.030)
Psychiatric Problem	a	-0.013***	(0.006)	-0.014**	(0.006)	-0.009	(0.008)
	b	-0.010***	(0.005)	-0.011**	(0.006)	-0.007	(0.006)
Longstanding Illness	a	0.189*	(0.072)	-0.183***	(0.048)	-0.343***	(0.051)
	b	0.220**	(0.057)	-0.120**	(0.052)	-0.272***	(0.057)
Unable to work	a	0.046***	(0.028)	-0.017**	(0.008)	-0.016*	(0.008)
	b	0.016*	(0.015)	-0.009***	(0.004)	-0.007	(0.004)

* p<0.10, ** p<0.05, *** p<0.01 - test of the underlying coefficient being 0

Notes: a: without objective health data, b: with objective health data. Reference category: SRH – fair. Marginal effects calculated as $\Phi(\bar{x}\hat{\beta} + \hat{\beta}_{SRH}) - \Phi(\bar{x}\hat{\beta})$ at the mean of all other variables, \bar{x} . Robust standard errors in parentheses. Objective health data: Blood analytes (haemoglobin, ferritin, total cholesterol, hdl cholesterol, c-reactive protein, fibrinogen), systolic and diastolic blood pressure, bmi, waist/hip-ratio. Included controls in all estimations: Age groups, income quintiles, dummies for education outcomes, race, employment status, marital status, self-reported conditions at baseline, functional limitations at baseline, unable to work at baseline, smoking status, parents cause of death (stroke or heart attack).

Table 7: Influence of Selection Mechanisms

Health Event		Missing data		Attrition	
		M.E.	SE	M.E.	SE
Dead (all causes)	m	-.02	.01		
	f	-.02**	.01		
Functional Limitation	m	-.02	.02	.01	.04
	f	-.02	.02	-.04	.03
Major Condition	m	-.04**	.02	.01	.03
	f	-.01	.01	.03	.02
Cancer	m	-.002	.01	.00	.01
	f	.00	.01	.03**	.01
Heart Attack	m	-.01	.01	.01	.01
Angina	m	-.02*	.01	.00	.01
	f	-.00	.01	-.01	.00
Stroke	m	-.01	.01	-.00	.00
	f	-.01	.01	.0007**	.0007
Diabetes	m	-.01	.01	.01*	.01
	f	.00	.01	-.00	.00
Arthritis	m	.02	.01	.03	.03
	f	.03**	.01	.03	.03
Psychiatric Problem	m	-.00	.07	.001*	.00
	f	.01	.01	.01	.01
Longstanding Illness	m	-.03	.02	-.01	.05
	f	-.00	.02	.02	.04
Unable to work	m	-.02***	.01	-.06	.07
	f	-.01	.01	-.13*	.06

* p<0.10, ** p<0.05, *** p<0.01 - test of the underlying coefficient being equal to 0.

Notes: m: male, f: female. Probit models with dependent variables in column 1. Missing data: Whole sample included, marginal effects of the indicator of availability of objective health data. Attrition: Sample with objective health data, marginal effect of indicator of attrition. Marginal effects calculated as $\Phi(\bar{x}\hat{\beta} + \hat{\beta}_I) - \Phi(\bar{x}\hat{\beta})$ at the mean of all other variables, \bar{x} . I denotes the indicator of interest. Robust standard errors of the marginal effect reported. Included controls in all estimations: Age groups, income quintiles, dummies for education outcomes, race, employment status, marital status, self-reported conditions at baseline, functional limitations at baseline, unable to work at baseline, smoking status, parents cause of death (stroke or heart attack).

Table 8: Weighting-Women

Health Event		SRH – very bad/bad		SRH – good		SRH – very good	
Dead (all causes)	a	0.037**	(0.023)	-0.020**	(0.009)	-0.022**	(0.009)
	b	0.067**	(0.038)	-0.019	(0.014)	-0.026*	(0.014)
Functional Limitation	a	0.182***	(0.072)	-0.135***	(0.030)	-0.179***	(0.031)
	b	0.114	(0.078)	-0.181***	(0.035)	-0.231***	(0.035)
	c	0.148**	(0.077)	-0.138***	(0.032)	-0.185***	(0.033)
Major Condition	a	-0.029	(0.030)	-0.021	(0.020)	-0.013	(0.023)
	b	-0.011	(0.045)	-0.026	(0.025)	-0.020	(0.027)
	c	-0.019	(0.035)	-0.018	(0.020)	-0.012	(0.023)
Cancer	a	-0.029***	(0.006)	-0.014	(0.012)	0.011	(0.014)
	b	-0.031***	(0.006)	-0.011	(0.012)	0.009	(0.015)
	c	-0.029***	(0.006)	-0.012	(0.012)	0.010	(0.014)
Angina	a	0.010	(0.018)	-0.014**	(0.007)	-0.021***	(0.007)
	b	0.051*	(0.040)	-0.022**	(0.011)	-0.033***	(0.010)
	c	0.011	(0.020)	-0.013*	(0.007)	-0.019**	(0.007)
Stroke	a	-0.001	(0.002)	0.001	(0.001)	-0.001	(0.002)
	b	-0.003	(0.002)	-0.0001	(0.002)	-0.001	(0.003)
	c	0.001	(0.004)	0.0003	(0.001)	-0.001	(0.002)
Diabetes	a	-0.002	(0.001)	-0.002	(0.002)	-0.001	(0.002)
	b	-0.003*	(0.001)	-0.003*	(0.002)	-0.002	(0.002)
	c	-0.003	(0.001)	-0.002	(0.002)	-0.002	(0.002)
Arthritis	a	0.025	(0.059)	-0.037	(0.028)	-0.054*	(0.029)
	b	0.014	(0.056)	-0.038	(0.028)	-0.058*	(0.028)
	c	0.004	(0.054)	-0.041	(0.028)	-0.058*	(0.029)
Psychiatric Problem	a	0.009	(0.025)	-0.029***	(0.011)	-0.046***	(0.011)
	b	0.008	(0.021)	-0.021**	(0.010)	-0.039***	(0.009)
	c	0.004	(0.022)	-0.024**	(0.011)	-0.043***	(0.010)
Longstanding Illness	a	0.211*	(0.072)	-0.089**	(0.044)	-0.223***	(0.049)
	b	0.194**	(0.044)	-0.085**	(0.041)	-0.202***	(0.048)
	c	0.212*	(0.068)	-0.100**	(0.044)	-0.225***	(0.049)
Unable to work	a	0.035***	(0.019)	-0.013***	(0.006)	-0.007	(0.005)
	b	0.020*	(0.015)	-0.015***	(0.006)	-0.008	(0.005)
	c	0.034***	(0.019)	-0.013***	(0.006)	-0.007	(0.005)

* p<0.10, ** p<0.05, *** p<0.01 - test of the underlying coefficient being 0

Notes: a: unweighted, b: weighted missing objective health data, c: weighted attrition. Marginal effects, robust standard errors in parentheses.

Table 9: Weighing-Men

Health Event		SRH – very bad/bad		SRH – good		SRH – very good	
Dead (all causes)	a	0.032	(0.034)	-0.018	(0.018)	-0.039**	(0.018)
	b	0.055	(0.052)	-0.039	(0.028)	-0.070	(0.027)
Functional Limitation	a	0.131*	(0.083)	-0.116***	(0.036)	-0.168***	(0.037)
	b	0.123	(0.087)	-0.133***	(0.039)	-0.187***	(0.039)
	c	0.130*	(0.086)	-0.113***	(0.038)	-0.163***	(0.039)
Major Condition	a	0.051	(0.056)	-0.070**	(0.028)	-0.078**	(0.029)
	b	0.071	(0.064)	-0.063*	(0.033)	-0.070**	(0.034)
	c	0.054	(0.058)	-0.074**	(0.029)	-0.083***	(0.030)
Cancer	a	0.010	(0.033)	0.001	(0.015)	0.003	(0.017)
	b	0.005	(0.029)	-0.001	(0.016)	0.003	(0.018)
	c	0.007	(0.030)	0.001	(0.015)	0.002	(0.017)
Heart Attack	a	0.003	(0.022)	-0.005	(0.012)	-0.013	(0.013)
	b	0.018	(0.034)	0.009	(0.018)	-0.005	(0.019)
	c	0.005	(0.022)	-0.003	(0.011)	-0.010	(0.012)
Angina	a	-0.003	(0.015)	-0.017	(0.011)	-0.015	(0.012)
	b	-0.006	(0.018)	-0.012	(0.014)	-0.005	(0.015)
	c	-0.011	(0.012)	-0.016	(0.011)	-0.013	(0.012)
Stroke	a	0.008	(0.011)	0.003	(0.004)	0.001	(0.005)
	b	-0.001	(0.006)	0.006	(0.007)	0.003	(0.008)
	c	0.010	(0.013)	0.001	(0.004)	-0.0002	(0.005)
Diabetes	a	0.047	(0.046)	-0.001	(0.016)	-0.006	(0.018)
	b	0.069*	(0.055)	0.005	(0.018)	-0.0002	(0.020)
	c	0.057	(0.050)	-0.002	(0.016)	-0.009	(0.018)
Arthritis	a	-0.053	(0.029)	-0.018	(0.027)	-0.039	(0.030)
	b	-0.051	(0.030)	-0.014	(0.028)	-0.027	(0.031)
	c	-0.045	(0.031)	-0.024	(0.028)	-0.038	(0.030)
Psychiatric Problem	a	-0.010***	(0.005)	-0.011**	(0.006)	-0.007	(0.006)
	b	-0.009***	(0.005)	-0.010**	(0.005)	-0.006	(0.005)
	c	-0.008***	(0.004)	-0.010**	(0.005)	-0.006	(0.005)
Longstanding Illness	a	0.220**	(0.057)	-0.120**	(0.052)	-0.272***	(0.057)
	b	0.197**	(0.054)	-0.117**	(0.050)	-0.260***	(0.057)
	c	0.215*	(0.063)	-0.146***	(0.053)	-0.302***	(0.058)
Unable to work	a	0.016*	(0.015)	-0.009***	(0.004)	-0.007	(0.004)
	b	0.012	(0.013)	-0.010**	(0.005)	-0.007	(0.004)
	c	0.014	(0.015)	-0.011***	(0.005)	-0.008*	(0.004)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ - test of the underlying coefficient being 0

Notes: a: unweighted, b: weighted missing objective health data, c: weighted attrition. Marginal effects, robust standard errors in parentheses.

Figures

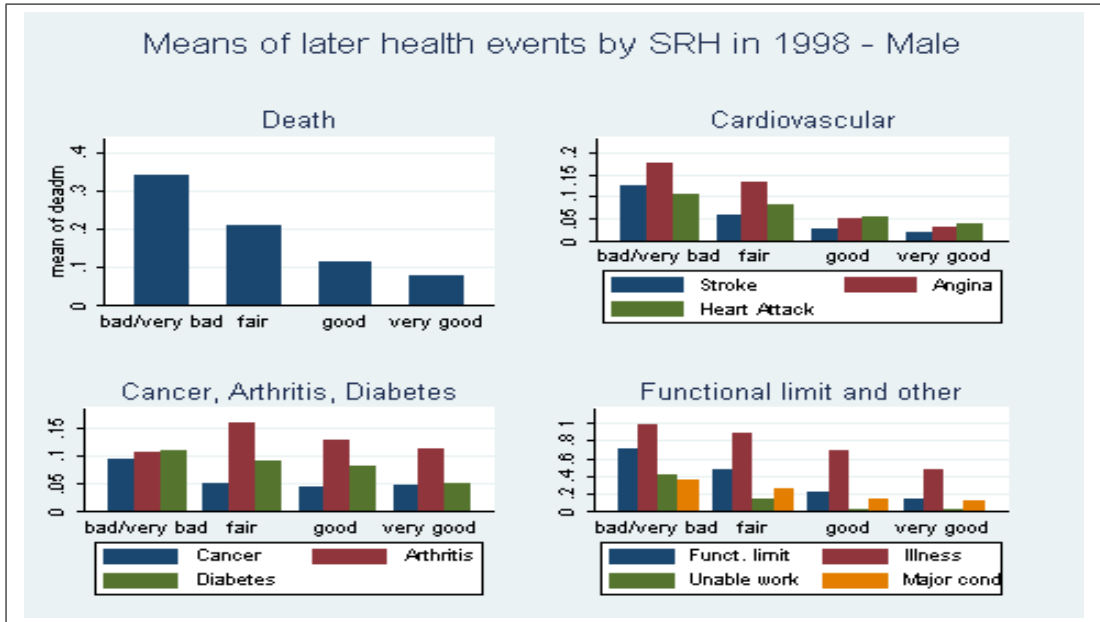


Figure 1: SRH and later health events - Men

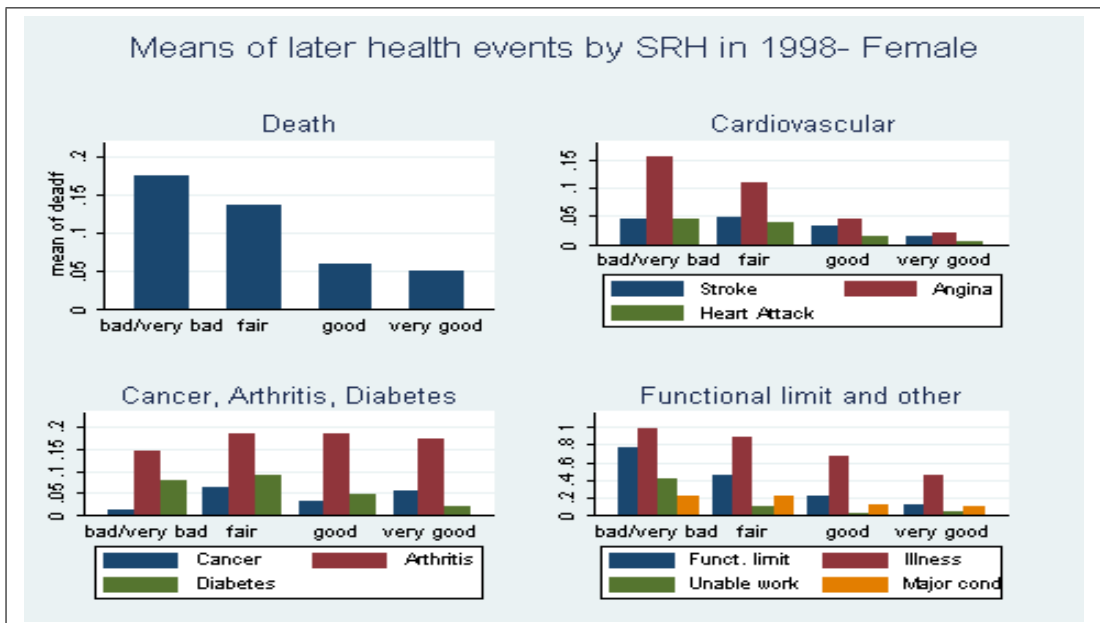


Figure 2: SRH and later health events - Women