

Wuppermann, Amelie C.

**Conference Paper**

## Do I know more than my body can tell? An Empirical Analysis of Private Information

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Economic Aspects of Health, No. B18-V4

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Wuppermann, Amelie C. (2010) : Do I know more than my body can tell? An Empirical Analysis of Private Information, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Economic Aspects of Health, No. B18-V4, Verein für Socialpolitik, Frankfurt a. M.

This Version is available at:

<https://hdl.handle.net/10419/37182>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Do I know more than my body can tell?

An Empirical Analysis of Private Information.

Amelie C. Wuppermann\*

Preliminary Version

July 31, 2010

## Abstract

Economic theory stresses the importance of adverse selection in competitive insurance markets. The empirical evidence for adverse selection in different health-related insurance markets, however, is mixed. This study evaluates whether different degrees of private information in different insurance markets can explain the conflicting evidence. The analysis uses data from the English Longitudinal Study of Ageing, one of the rare longitudinal data sets with objectively measured health information. Using self-rated health as a proxy for private information this study finds that extensive information collection by the insurer can eliminate private information on health risks and thus the scope for adverse selection in life and health insurance.

JEL-Code: D82, I10, C33

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

---

This research has profited from comments by seminar participants at the University of Munich, at the ifo Institute, at Tilburg University, at the ASHecon Conference in Cornell, at the ECHE in Helsinki, and at the SMYE in Istanbul. I received insightful comments from Luc Bissonnette, Florian Heiß, Hendrik Jürges, Helmut Farbmacher, Daniel McFadden, Erik Meijer and Joachim Winter. Financial support from the Deutsche Forschungsgemeinschaft through GRK 801 is gratefully acknowledged.

\*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 about risk types in competitive insurance markets are known to induce inefficient outcomes due to adverse selection. In particular, the standard textbook model by Rothschild and Stiglitz (1976) and its extensions robustly predict that higher risk individuals buy more insurance coverage than lower risk individuals (Chiappori et al. (2006)). The empirical evidence on this matter, however, varies for different insurance markets (Cohen and Siegelman (2010)).

Multiple dimensions of asymmetric information, for example private information on risk preferences in addition to private information on risk type, have been stated as a reason for the mixed evidence (De Meza and Webb (2001), Cutler et al. (2008)). The absence of useful private information and thus the absence of scope for adverse selection in some but not in other insurance markets could provide a different explanation for the mixed findings (Cohen and Siegelman (2010)). In this paper, I evaluate empirically whether different degrees of private information exist in different insurance markets.

The focus of this paper lies in private information in the markets for health insurance, life insurance and annuities. The evidence of adverse selection varies between these markets. While life insurance markets are not found to be adversely selected, annuities markets show signs of adverse selection even though the insured risk in both markets is related to the timing of death. Similarly, different health insurance markets show different results for adverse selection. Differences in private information could explain these findings because insurance companies in the different markets collect and use different types of information for insurance underwriting, i.e. for risk classification and calculation of premia.<sup>1</sup>

The analysis is conducted using data from the English Longitudinal Study of Ageing (ELSA). The ELSA data set contains a broad range of health measures that are typically

---

<sup>1</sup>Possible explanations for use of different information for underwriting in markets that insure similar risks could be legal restrictions, political economy concerns (Finkelstein and Poterba (2006)), specificities of the market structure as put forward by Kesternich and Schumacher (2009), or different demand for underwriting in the different markets (Browne and Kamiya (2009)).

collected by insurance companies. In addition to self-reported pre-existing health conditions, ELSA is one of the few longitudinal data sets that provides health data which are objectively measured and reported by a nurse: Results of a blood sample analysis, blood pressure measurement, objectively measured body mass index (BMI), and waist-hip-ratio are available.

Self-rated general health (SRH) on a 5-point scale from very bad to very good serves as a proxy for private information on health risks. It was chosen for two reasons: First, a large number of studies have shown that SRH contains information on actual health and future health events like death.<sup>2</sup> Second, SRH is a non-verifiable measure in the sense that insurance companies have no means to verify whether an individual's statement of SRH is true. This is in contrast to other self-reported measures like an individual's comorbidities or family health history which can be verified by going back to health records. Its non-verifiability makes SRH particularly valuable for analyzing the existence of private information. Information about health risks that is only contained in SRH necessarily remains private. In this study it is therefore interpreted as evidence for the existence of private information and thus scope for adverse selection when SRH contains information about subsequent health events that is additional to the information in verifiable measures.

As I am interested in private information in the markets for life insurance, annuities and health insurance, I analyze the information in SRH for whether an individual dies in the next years and for whether an individual is newly diagnosed or has a recurrence of a major health condition. The results indicate that SRH contains information on dying or surviving the next 10 years and on being diagnosed with a costly major health condition in the next 8 years, when only a limited number of additional control variables is included in the analysis. With the inclusion of medical information and in particular with the inclusion of objectively measured health data, however, neither dying or surviving nor being diagnosed with a major condition in the future is significantly predicted by SRH.

These results in combination with different use of information for underwriting in differ-

---

<sup>2</sup>For overviews on studies analyzing the relationship between SRH and subsequent death see Idler and Benyamini (1997), Benyamini and Idler (1999) and DeSalvo et al. (2006).

ent insurance markets could explain the mixed evidence of adverse selection in life insurance and annuities markets, and in different health insurance markets. While in life insurance markets stringent underwriting is performed and no evidence is found for adverse selection, in annuities markets only limited information is used for underwriting and evidence for adverse selection is found. Similarly, no individual underwriting is performed in group health insurance markets and the markets show evidence for adverse selection, while in individual health insurance stringent underwriting is performed and no adverse selection is found.

The remainder of the paper is structured as follows. Section 2 gives an overview on the empirical evidence for private information in health-related insurance markets, Section 3 introduces the data used for the analysis. Section 4 outlines the estimation strategy, results are shown in Section 5, Section 6 reports the results of robustness analyses, and the last section concludes.

## **2 Literature on health-related private information**

Cohen and Siegelman (2010) provide an up-to-date overview on empirical studies of adverse selection in different insurance markets. The health-related insurance markets featured in their summary are the markets for life insurance, annuities, health insurance and long-term care.

While in markets for annuities evidence for adverse selection is found (Finkelstein and Poterba (2002), (2004), (2006) and McCarthy and Mitchell (2010)), there is only very little evidence for adverse selection in life insurance markets. He (2009) finds evidence that individuals who newly buy life insurance die earlier than individuals who do not buy life insurance when controlling for some variables that insurance companies use for risk classification and calculation of premia. This study, however, might not sufficiently control for risk classification undertaken by the insurers as objectively measured health data is not available. Furthermore, Cawley and Philipson (1999), Hendel and Lizzeri (2003) and McCarthy and Mitchell (2010) find no evidence for adverse selection into life insurance.

Hendel and Lizzeri (2003) point out that the different findings in life insurance and annuities markets might be due to differences in underwriting in the two markets.

The evidence for adverse selection in markets for health insurance is mixed. Especially the US employer-sponsored health insurance market is found to suffer from adverse selection as the studies summarized by Cutler and Zeckhauser (2000) show. A more recent study that finds evidence for adverse selection in employer-sponsored health insurance is Cutler et al. (2009). The authors study individuals' choices between different health plans offered by their employer and find that individuals who switch to more generous plans are older and can expect higher health-care costs in the future.

Other health insurance markets, however, are often not found to suffer from adverse selection. Buchmueller et al. (2004) show that sicker individuals are not more likely to buy private supplementary health insurance in France, Propper (1989) and Doiron et al. (2008) find similar results for the British and Australian private health insurance markets, respectively. Furthermore, the US market for Medigap coverage is not found to be affected by adverse selection (Fang et al. (2008)).

Like in the case of life insurance and annuities, the use of different information for underwriting could provide a possible explanation for different degrees of adverse selection in different health insurance markets. While there is typically no underwriting in employer-sponsored insurance, buying health insurance on an individual level requires the applicant to disclose certain type of relevant information to the insurance company that is used for risk classification and calculation of premia (see for example Mossialos and Thomson (2009)).

The stringency of the underwriting process is not the reason for the absence of adverse selection in the US market for long-term care. Finkelstein and McGarry (2006) show that individuals hold private information about their risk of needing long-term care even when considering the insurer's risk classification. Nevertheless, the market of long term care does not show a positive correlation between risk and insurance status. The authors argue that heterogeneity in risk preferences could result in advantageous selection that outweighs

adverse selection. Individuals who are risk-averse are more likely to buy insurance. They are also typically of lower risk because they engage in risk lowering activities. Taken together this results in a negative correlation between risk type and insurance status. With an increasing availability of genetic information on disease risks, however, the degree of private information about future need of long term care is likely to increase. As Oster et al. (forthcoming) show using information on current tests for Huntington disease this might result in adverse selection in long-term care markets in the future.

Private information in markets related to health risks is analyzed by Banks et al. (2007). Similar to my analysis, the authors use SRH as a proxy for private knowledge and analyze the onset of different future health events in a Canadian data set. 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 women but not for men when controlling for pre-existing conditions, socio-economic variables and some risk factors (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. As objective health data is not included in the data set, the authors cannot fully investigate the effects of medical underwriting on the degree of private knowledge.

My analysis expands on the literature in two ways. First, it sheds light on the question whether different degrees of private information in different insurance markets can explain the mixed empirical findings on adverse selection in these markets. Second, the availability of a rich data set that includes objectively measured health information allows me to analyze the difference that medical underwriting makes for the scope for adverse selection.

### **3 Data**

The English Longitudinal Study of Ageing (ELSA) is a rich panel data set which contains socio-demographic, economic and health-related data for 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 from three ELSA waves collected in 2002, 2004 and 2006 is used.<sup>3</sup>

While ELSA was highly influenced by and modelled on the US Health and Retirement Study (HRS), the design of ELSA differs from that of the HRS in one important feature: In addition to the biannual interview, every four years a nurse visit is conducted as part of ELSA. Due to this nurse visit objectively measured blood pressure, results of a blood sample analysis, and anthropometric data are available. Up to now, data on nurse visits is available for waves 0 and 2. In wave 0, a blood sample analysis was only carried out for individuals in the 1998 HSE year. As a focus of this study is the scope for adverse selection when insurance companies are allowed to collect and use outcomes of medical screening, the analysis is conducted using only ELSA sample members that were sampled for the 1998 HSE.

The ELSA data in wave 0 contains 8,267 individuals from HSE 1998 (7,807 core sample members and 459 younger partners). Everyone who conducted the interview was eligible for the nurse visit in HSE, therefore core sample members and younger partners from wave 0 are included in this analysis. As Table 1 shows, the data contains information on age, gender, race, social occupational class, marital status and smoking status for nearly all individuals in the sample. Individuals older than 90 are not included in the analysis as their exact age is not known. This reduces the sample size by 1 percent.

The medical information available is displayed in Tables 2 and 3. As columns 2 (women) and 5 (men) of Table 2 show, self-reported health data and information about co-morbidities are also broadly available. Objectively measured health data, however, is missing for a relatively large share of individuals. Overall, all health measures in wave 0 are available for about 50% of the sample.

---

<sup>3</sup>For a more thorough description of the ELSA data see Marmot et al. (2009).



The sample size shrinks further due to attrition when analyzing a new diagnosis or recurrence of a major health condition. Information on this dependent variable is derived from an individual's answers to questions in subsequent waves of ELSA on whether they have been diagnosed with cancer, heart disease and/or stroke. The information is thus only available if the individual appears again in ELSA after wave 0. Overall, of the 4,305 individuals 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.

Information on death, however, is collected regardless of attrition 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.

As can be inferred from the differences in the means between columns 1 and 3 (women) and between columns 4 and 6 (men) of Tables 1 and 2 the individuals who remain in the sample when item non-response in wave 0 and attrition are taken into account differ in many aspects from the overall sample. They are on average younger, more likely to be white or married and less likely to be smokers. Also, they are healthier in terms of both, subjective and objective health measures. If the predictive power of SRH for subsequent health events varies with health or age, ignoring the two selection mechanisms might result in biased estimates. Furthermore, the existence of unobservable influences on selection that also affect the future health outcomes would lead to inconsistent estimates when ignoring selection. Inverse probability weighting is employed to correct for the two selection mechanisms.

Figures 1 and 2 present a first glance at the relationship between SRH and the future 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 year 2008. For men, a similar graded relationship can be observed between SRH and a diagnosis of a major condition within the next 8 years. Similarly, less women in very good or good health are diagnosed with a major condition within the next 8 years

compared to women in fair, bad or very bad health. The graded relationship does not hold, however, for very bad or bad health compared to fair health for women.

## 4 Estimation Strategy

The existence of private information in health insurance, life insurance and annuities markets is investigated by regressing indicator variables for the occurrence of future health events on categories of SRH and different control variables at a baseline year. To capture the risk that is insured in life insurance and annuities markets the dependent variable is whether an individual is dead 10 years after the baseline interview. The insured risk in health insurance markets is captured by a variable that indicates whether an individual is newly diagnosed or has a recurrence of heart disease, cancer or stroke within 8 years after the initial interview. These conditions belong to the most costly conditions at the per capita level (Druss et al. (2002)) and are thus good proxies for the risk insured in health insurance markets, namely high medical expenses.

The information about the future health events that is contained in SRH is interpreted as evidence for private information. SRH might, of course, not capture all private knowledge on health and mortality risk as SRH asks about health at the time of the interview and not about expected changes in health or death in the future. A better suited proxy for private information might be subjective life expectancy. Hurd and McGarry (2002) find that subjective life expectancy contains an expectational component in addition to the health component that is also captured by SRH. Subjective life expectancy, however, is only elicited starting from ELSA wave 1. Objectively measured health information is only available in wave 0 and 2, thus using subjective life expectancy in the analysis would reduce follow-up time significantly. Subjective life expectancy is, however, included in the analysis as a robustness check.

Let  $y_i^{j*}$ ,  $j \in \{D, M\}$  denote latent variables for the future health events, where  $D$  stands for death and  $M$  for being diagnosed with a major condition, for individual  $i$ . Each

of the  $y_i^{j*}$ 's can be represented by the following equations

$$\begin{aligned} y_i^{j*} &= \beta_0^j + \beta_1^j SRH_{1i} + \beta_2^j SRH_{2i} + \beta_3^j SRH_{3i} + \beta_4^j X_{ai} + \epsilon_i^j \\ y_i^j &= \mathbb{I}(y_i^{j*} > 0) \end{aligned} \quad (1)$$

where the  $SRH$  variables represent dummies for the three categories, very bad/bad, good and very good SRH with fair SRH as the reference category.  $X_a$  is a vector of variables that represent the information that insurance companies in insurance market  $a$  collect and use for underwriting.

$\epsilon_i^j$  captures unobservables influences on the latent future health event  $j$ . Under the assumptions that  $\epsilon_i^j \sim N(0, 1)$  and that the correlation between  $\epsilon^D$  and  $\epsilon^M$ ,  $\rho_{DM}$ , is equal to 0, I estimate  $Pr(y_i^j = 1)$  for each of the health events independently using single equation probit models.<sup>4</sup>

Table 4 displays the variables that are typically collected and used in the application process in different insurance markets in the UK and in the US. The listed variables represent information that is used for risk classification and calculation of premia. As has been noted by Finkelstein and Poterba (2006), insurance companies may have additional information about applicants that is not used in underwriting. These “unused observables” do not mitigate the scope for adverse selection and are thus not included as controls in  $X$ .

The vector of control variables for the investigation of private information on the insured risk in group health insurance markets,  $X_{GroupHI}$ , does not include any variables as there is typically no individual underwriting in these insurance markets. In annuities markets, underwriting is based on the individual’s age and sex. As I estimate separate models for men and women the vector of control variables in annuities markets,  $X_{Annuities}$ , includes only linear splines in age.

Considerably more information is used for underwriting in life and health insurances in both, the UK and the US. In addition to age and sex, medical information of the individual

---

<sup>4</sup>The assumption of independence between the error terms can be relaxed and a bivariate equation probit models be estimated. For this analysis, however, the assumption of independent error terms seems appropriate as only private information on the particular risk insured in each specific market, i.e. either the risk of dying/surviving or the risk of high medical expenses, is of interest.

is typically used. To capture the medical information I include self-reported conditions, linear splines of objectively measured BMI, waist-to-hip ratio, and blood values in  $X_{Life}$  and  $X_{IndividualHI}$ . Information on whether individuals smoke or used to smoke is typically also used in health and life insurance underwriting and is thus included as control variable. Furthermore, life insurance and some health insurance companies use medical history of an individual's family which is proxied for by parents' cause of death in my analysis.

Insurance companies sometimes use information on occupational status, dangerous occupations, hazardous hobbies, risky travel destinations, residence/citizenship, and alcohol or drug abuse for underwriting. As a proxy for dangerous occupation, I include dummies for occupational social class. The other variables are unfortunately not well captured in the ELSA data.

## 5 Results

Average marginal effects after probit estimation of equation 1 are reported in Tables 5-8. The first two tables show the information in SRH on whether an individual dies within the next 10 years, for men and women respectively. Tables 7 and 8 display average marginal effects for a new diagnosis or recurrence of one of the major health conditions, heart disease, cancer or stroke, in the next 8 years. All tables show results that are weighted to correct for missing data.<sup>5</sup>

The first columns of Tables 5 and 6 show that there is significant information in SRH in 1998 for whether an individual is known to be dead 10 years later when no further variables are added as controls. As in annuities markets, age and sex are typically used for underwriting, column 2 of each of the tables is of specific interest. For both genders, all three SRH coefficients are highly significantly different from 0. The average marginal effects indicate that conditional on age women (men) in very good SRH in 1998 are 8 (13) p.p. less likely to be dead in 2008 than women (men) who rate their health as fair.

---

<sup>5</sup>The estimation of the weights is described in the appendix. Unweighted results do not differ significantly from the weighted results.

When interpreting conditional information in SRH as private information these results show evidence for a scope for adverse selection in annuities markets.

From left to right more control variables are added to the estimation. The last columns of Tables 5 and 6 report average marginal effects when age, smoking information, medical conditions, family health history, social occupational class and objective health data are included as controls. The p-values of the Wald test for joint significance indicate that for both genders the three SRH coefficients are still jointly significantly different from 0 at a 10 percent significance level. The inclusion of the additional controls, however, and in particular the inclusion of the objectively measured health information, leads to an attenuation in the average marginal effects and to a reduction in the significance of the underlying coefficients of SRH. Thorough underwriting, and in particular medical underwriting that includes blood tests and other objectively measured health data, is thus able to significantly reduce private information about mortality risks.

Tables 7 and 8 show similar results for the diagnosis or re-diagnosis of a major condition in 8 years after the baseline interview. With no additional controls added, the coefficient of the three SRH categories are jointly significantly different from zero for women and men. When there is no individual underwriting, as in the case of employer-sponsored health insurance, there thus seems to be private information and therefore scope for adverse selection.

The inclusion of additional controls from left to right in Tables 7 and 8 results in a reduction in the information in SRH. For both genders, the inclusion of self-reported medical conditions and the number of prescription drugs taken in column 4 leaves no additional explanatory power in SRH for the diagnosis or re-diagnosis of a major health condition. Taking information in SRH on a future diagnosis of a major condition as private information that is relevant for the insured risk in health insurance markets, the results indicate that medical underwriting might also be a crucial determinant of the degree of private information in private health insurance markets.

Overall, the results help to strengthen the intuition that it is important to consider the

exact information that insurance companies use for underwriting when analyzing adverse selection. For example, ignoring the fact that in some insurance markets objective health data is available to the insurance company likely overstates the scope for adverse selection. It should be born in mind, however, that SRH is only a proxy for private information about the different health events. It might be the case that individuals have private knowledge on their risk of dying or contracting a specific disease that might just not be reflected in SRH. It would thus be interesting to elicit individuals' knowledge on specific future health events by asking more specific questions.

## 6 Robustness

In this section, I present different sensitivity analyses. First, subjective life expectancy is included in the analysis instead of SRH as it might be a better proxy for private information on mortality risks. Second, the analysis is conducted for younger individuals only as some insurance companies set maximum ages at which individuals can buy new policies (Mossialos and Thomson (2009)). Third, the time-horizon of the dependent variables is shortened to 4 years to reflect the use of short-term contracts that is especially common in private health insurance markets.

Table 9 displays the results using subjective life expectancy instead of SRH as proxy for private information.<sup>6</sup> This variable is only available in ELSA starting from wave 1. The earliest wave that contains both, subjective life expectancy and objectively measured health data, is wave 2 and I thus use 2006 as baseline year for this analysis. The dependent variable is whether an individual is known to be dead by 2008.<sup>7</sup> For means of comparison, I also report results using SRH in 2006.

The first panel in Table 9 reports average marginal effects for 3 categories of subjective life expectancy after probit estimation. Significant information in subjective life expectancy

---

<sup>6</sup>Subjective life expectancy in ELSA is elicited with the question "What are the chances that you will live to be x or more?" Where x depends on the individual's age at the time of the survey. The average time horizon for this question in the estimation sample is 15 years.

<sup>7</sup>Information for later years is not available up to now.

is found for both genders when only age splines are included as controls. Individuals who report a 50-74 or 75-100 percent chance of surviving the next years are less likely to die within the next two years than individuals who report a 25-49 percent chance. For both genders, the underlying coefficients remain jointly statistically significant when the full set of available life insurance controls in ELSA wave 2 is included as control variables. The average marginal effects, however, lose statistical significance with inclusion of the additional controls. There is some evidence for private information in subjective life expectancy on dying within the next two years. However, the time horizon in this robustness analysis is different from the main analysis and the life insurance controls are not fully identical with the controls in the main analysis because different variables are contained in the different ELSA waves. The results are thus not directly comparable with each other.

Whether subjective life expectancy is a better proxy for information on mortality risk can be seen by comparing the results reported in the first and the second panel of Table 9. For women, subjective life expectancy seems to contain more information on dying in the next years than SRH. While the underlying coefficients of the former remain jointly statistically significant at all common significance levels after the inclusion of the life insurance controls, the latter only remain jointly significant at a significance level of 15 percent. For men, the difference between using SRH and subjective life expectancy is less pronounced. The underlying coefficients of both variables remain highly jointly significant after the inclusion of the life insurance controls. Using SRH and subjective life expectancy would thus result in the same conclusion about the existence of private information. There is only weak evidence that subjective life expectancy is indeed a better proxy for private information on mortality risk than SRH.

The results for the second and third robustness analyses are summarized in Table 10. The first two panels display average marginal effect of the three SRH categories for women and men younger than 65. The first two columns report the average marginal effects when the dependent variable is whether an individual is known to be dead by 2008 or not.

Column 3 and 4 report results for whether an individual is diagnosed or re-diagnosed with a major health condition as dependent variable.

With only age as controls, there is significant information in SRH on dying in the next 10 years for men and women younger than 65. Including the full set of life insurance underwriting controls eliminates all private information captured in SRH. Similarly, SRH contains significant information for diagnoses or re-diagnosis of a major condition when no controls are included. When including the full set of private health insurance controls some significant information remains in the coefficients of SRH. However, the average marginal effects are never significant in the full control setting. Limiting the possible ages at which individuals can buy individual private health or life insurance thus does not seem to contradict the conclusion that thorough underwriting has the potential to eliminate private information that exists a priori on health and mortality risks.

Panels 3 and 4 of Table 10 report the average marginal effects of the three SRH categories when the dependent variable is whether an individual is known to be dead by ELSA wave 1 (columns 1 and 2) and whether an individual reports a diagnosis or recurrence of a major health condition in wave 1 (columns 3 and 4). Similarly to the results of the main analysis, there is information in SRH on death with limited controls that diminishes when the full set of life insurance controls is included.

For women, the results are not changed qualitatively when analyzing the information in SRH for a diagnosis or recurrence of a major health condition. There is private information in SRH when including no controls that disappears with inclusion of the full set of health insurance control variables. For men, some significant information remains in SRH for the diagnosis or recurrence of a major condition even after the full set of health insurance controls is included in the analysis. The underlying coefficient of very good SRH is significant at the 5 % level. The average marginal effect, however, is reduced by almost a factor of 1/2 compared to the estimation with no controls and is no longer statistically significantly different from 0.

Overall, the robustness analyses do not contradict the main conclusion that different



degrees of private information remain after underwriting in different insurance markets. While there is some evidence that subjective life expectancy is a better proxy for information on mortality risk than SRH, the reduction in information contained in the former variable is similar to the reduction in the information contained in SRH when additional controls are included. Furthermore, private information on both, mortality and health risk, is reduced through underwriting for younger individuals and when the realization of the risk is calculated at a shorter term.

## 7 Conclusion

The mixed empirical evidence on adverse selection in different insurance markets stands in contrast to the robust predictions on the existence of adverse selection in theoretical models. This divergence has engendered research that tries to reconcile theoretical and empirical results. This paper focuses on one possible explanation for the mixed findings in the empirical literature: differences in the scope for adverse selection in different insurance markets. In insurance markets in which insurance companies collect a lot of information about their applicants and use this information for underwriting, private information might not exist and thus there might be no scope for adverse selection.

In this study, I focus on the markets for life insurance, annuities and health insurance. Private information in these markets is detected by using information in SRH on dying in the next years and on future diagnoses of diseases. Different verifiable measures are included as controls in addition to SRH to imitate different types of information used in underwriting in the different insurance markets. The information that remains in SRH when including the different sets of underwriting controls is interpreted as private information in the different insurance markets.

The analysis employs data from ELSA, one of the rare longitudinal data sets that provides objectively measured health data. These data allow to mimic medical underwriting with greater precision than is typically possible with data from population surveys that only include self-reported health information. As thorough medical underwriting that in-

cludes outcomes of medical exams is typically conducted in life insurance markets and private individual health insurance markets, ELSA's objectively measured health data is particularly valuable for my analysis.

Dying within the next 10 years is significantly related to SRH at baseline when only information on age and sex is additionally included in the analysis. When medical information and objectively measured health data is additionally included, however, the predictive power of SRH for death or survival vanishes. Similarly, there is significant information in SRH for the diagnosis or recurrence of a major health condition in the next 8 years with only limited controls. The inclusion of medical information reduces the amount of information contained in SRH to insignificant levels. These results are robust to changes in the time horizon and to restricting the sample to younger individuals.

There is thus evidence for private information on mortality risk and health risk if insurance companies do not use medical information for underwriting. When medical underwriting is conducted, however, the private information and thus the scope for adverse selection is eliminated.

These findings can potentially reconcile differences in evidence of adverse selection that are found between life insurance and annuities markets and between group and individual private health insurance markets. Only limited information is used for underwriting in group health insurance and annuities and these markets are found to suffer from adverse selection. In individual health and life insurance markets medical underwriting is typically performed and there is no evidence for adverse selection. My results strengthen the intuition that different degrees of private information exist in the different markets and can thus help to explain the different results for adverse selection.

## References

- Anon. (1999). Risk Classification in Individually Purchased Voluntary Medical Expense Insurance. The American Academy of Actuaries Issue Paper.
- Banks, J., Crossley, T., and Goshev, S. (2007). Looking for Private Information in Self-assessed Health. HEDG Working Paper, University of York 9.
- 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.
- Browne, M. and Kamiya, S. (2009). A Theory of the Demand for Underwriting. *Working Paper University of Wisconsin-Madison*.
- Buchmueller, T., Couffinhal, A., Grignon, M., and Perronnin, M. (2004). Access to physician services: does supplemental insurance matter? Evidence from France. *Health economics*, 13(7):669–687.
- Cawley, J. and Philipson, T. (1999). An empirical examination of information barriers to trade in insurance. *The American Economic Review*, 89(4):827–846.
- 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.
- Cohen, A. and Siegelman, P. (2010). Testing for Adverse Selection in Insurance Markets. *The Journal of Risk and Insurance*, 77(1):39–84.
- Cutler, D., Finkelstein, A., McGarry, K., and Center, L. (2008). Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance. NBER Working Paper 13746.
- Cutler, D., Lincoln, B., and Zeckhauser, R. (2009). Selection Stories: Understanding Movement Across Health Plans. NBER Working Paper 15164.

- Cutler, D. and Zeckhauser, R. (2000). The Anatomy of Health Insurance. In Culyer, A. and Newhouse, J., editors, *Handbook of Health Economics*, volume I. Amsterdam: Elsevier.
- De Meza, D. and Webb, D. (2001). Advantageous selection in insurance markets. *RAND Journal of Economics*, 32(2):249–262.
- DeSalvo, K., Bloser, N., Reynolds, K., He, J., and Muntner, P. (2006). Mortality Prediction with a Single General Self-Rated Health Question - 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.
- Druss, B., Marcus, S., Olfson, M., and Pincus, H. (2002). The most expensive medical conditions in America. *Health Affairs*, 21(4):105–111.
- 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. *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.

- He, D. (2009). The life insurance market: Asymmetric information revisited. *Journal of Public Economics*, 93(9-10):1090–1097.
- Hendel, I. and Lizzeri, A. (2003). The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance. *Quarterly Journal of Economics*, 118(1):299–327.
- Hurd, M. and McGarry, K. (2002). The Predictive Validity of Subjective Probabilities of Survival. *The Economic Journal*, 112(482):966–985.
- 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.
- Kesternich, I. and Schumacher, H. (2009). On the use of information in repeated insurance markets. Discussion paper, SFB/TR 15 Governance and the Efficiency of Economic Systems.
- Marmot, M., Banks, J., Nazroo, J., and Blundell, R. (2009). English longitudinal study of ageing: Wave 0 (1998, 1999 and 2001) and waves 1-3 (2002-2007). 11th Edition. Colchester, Essex: UK Data Archive [distributor] SN: 5050.
- McCarthy, D. and Mitchell, O. (2010). International Adverse Selection in Life Insurance and Annuities. In Tuljapurkar, S., Chu, C., Gauthier, A., Ogawa, N., and Pool, I., editors, *Riding the Age Wave*, volume 3, chapter 6. Springer.
- Mossialos, E. and Thomson, S. (2009). Private Health Insurance in the European Union. Brussels, European Commission (Directorate General for Employment, Social Affairs and Equal Opportunities).
- Oliveira, C. (2008). Nurse Visit - ELSA user day. Presentation at ELSA user day 2008, Economic and Social Data Service.
- Oster, E., Shoulson, I., Quaid, K., and Dorsey, E. (forthcoming). Genetic Adverse Selection: Evidence from Long-Term Care Insurance and Huntington Disease. *Journal of Public Economics*.

Propper, C. (1989). An econometric analysis of the demand for private health insurance in England and Wales. *Applied Economics*, 21(6):777–792.

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.

Wilson, C. (1977). A Model of Insurance Markets with Incomplete Information. *Journal of Economic Theory*, 16(2):167–207.

Table 1: Descriptives- ELSA Wave 0

Variable	Women			Men		
	A Mean	(% missing)	B Mean	A Mean	(% missing)	B Mean
<b>Age</b>						
Age < 50	.2	(0.00)	.23	.15	(0.00)	.18
Age 50-59	.29	(0.00)	.34	.31	(0.00)	.34
Age 60-69	.24	(0.00)	.25	.27	(0.00)	.30
Age 70-79	.19	(0.00)	.14	.19	(0.00)	.15
Age 80-89	.09	(0.00)	.04	.07	(0.00)	.04
<b>Occupational Class</b>						
Professional	.05	(0.00)	.07	.07	(0.00)	.09
Managerial -technical	.29	(0.00)	.32	.31	(0.00)	.33
Skilled - non manual	.16	(0.00)	.14	.09	(0.00)	.09
Skilled-manual	.25	(0.00)	.27	.33	(0.00)	.32
Semi-skilled manual	.16	(0.00)	.14	.14	(0.00)	.13
Unskilled manual	.07	(0.00)	.05	.05	(0.00)	.04
Other social class	.03	(0.00)	.01	.01	(0.00)	.01
<b>Marital Status</b>						
Married	.62	(0.00)	.70	.77	(0.00)	.80
Widowed	.22	(0.00)	.15	.08	(0.00)	.06
Separated/divorced/single	.16	(0.00)	.16	.15	(0.00)	.14
<b>Activity Status</b>						
Retired	.38	(0.00)	.32	.41	(0.00)	.36
Unemployed	.01	(0.00)	.02	.03	(0.00)	.02
Sick	.04	(0.00)	.02	.08	(0.00)	.06
Working	.36	(0.00)	.45	.47	(0.00)	.54
Inactive	.21	(0.00)	.19	.02	(0.00)	.01
<b>Race</b>						
White	.97	(0.00)	.98	.97	(0.00)	.98
<b>Smoking Behavior</b>						
Current Smoker	.22	(0.00)	.19	.22	(0.00)	.16
Ever Smoker	.56	(0.00)	.54	.74	(0.00)	.71
N	4,648		1,574	3,556		1,242

*Notes:*

A - Entire sample

B - Part of sample with no item non-response and no attrition

Table 2: Health Measures in ELSA Wave 0

Variable	Women			Men		
	A Mean	(% missing)	B Mean	A Mean	(% missing)	B Mean
<b>Self-Rated Health</b>						
Very bad/bad	.1	(0.00)	.06	.11	(0.00)	.06
Fair	.25	(0.00)	.21	.23	(0.00)	.19
Good	.39	(0.00)	.42	.37	(0.00)	.40
Very good	.27	(0.00)	.32	.29	(0.00)	.35
<b>Medical Conditions</b>						
Hypertension	.3	(0.11)	.23	.29	(0.03)	.24
Diabetes	.04	(0.02)	.02	.06	(0.00)	.04
Stroke	.03	(0.00)	.02	.04	(0.03)	.02
Heart Attack	.03	(0.02)	.01	.08	(0.03)	.04
Angina	.07	(0.00)	.04	.1	(0.03)	.05
Heart Murmur	.04	(0.02)	.04	.04	(0.00)	.02
Irregular Heart Rhythm	.07	(0.00)	.05	.08	(0.00)	.05
Oth Heart Problems	.02	(0.02)	.01	.03	(0.00)	.01
GHQ12 <sup>1</sup>	1.61	(6.26)	1.38	1.23	(6.21)	.92
No Longstanding Illness (LI) <sup>2</sup>	.45	(0.04)	.51	.43	(0.06)	.48
Non-limiting LI	.2	(0.04)	.21	.22	(0.06)	.24
Limiting LI	.36	(0.04)	.28	.36	(0.06)	.29
# Prescription drugs taken <sup>3</sup>	1.85	(14.44)	1.31	1.62	(12.77)	1.1
<b>Use of Medical Services</b>						
# of GP visits last 2 weeks <sup>4</sup>	.23	(0.02)	.2	.2	(0.03)	.16
# Hospital nights last year <sup>5</sup>	1.03	(0.04)	.45	1.15	(0.08)	.51
<b>Family Health History</b>						
Father dead	.86	(2.32)	.83	.89	(2.53)	.88
Mother dead	.73	(1.61)	.67	.76	(1.77)	.71
At least one parent died of						
Hypertension	.01	(1.25)	.01	.01	(1.88)	.01
Angina	.03	(1.25)	.02	.02	(1.88)	.03
Heart Attack	.27	(1.25)	.27	.28	(1.88)	.27
Oth Heart Problem	.14	(1.25)	.13	.15	(1.88)	.15
Stroke	.09	(1.25)	.09	.08	(1.88)	.09
Diabetes	.02	(1.25)	.03	.02	(1.88)	.02
<b>Objective Health Data<sup>6</sup></b>						
Haemoglobin<13 <sup>a</sup> , 11.5 <sup>b</sup> g/dL	.05	(32.66)	.03	.08	(27.33)	.06
Haemoglobin>18 <sup>a</sup> , 16.5 <sup>b</sup> g/dL	.002	(32.66)	.001	.003	(27.33)	.001
Ferritin< 25 <sup>a</sup> , 20 <sup>b</sup> µg/L	.14	(33.91)	.14	.06	(28.4)	.07
Ferritin>400 <sup>a</sup> , 200 <sup>b</sup> µg/L	.03	(33.91)	.02	.02	(28.4)	.02
Total cholesterol>5 mmol/L	.8	(33.54)	.79	.75	(29.67)	.77
HDL cholesterol<1 <sup>a</sup> , 1.2 <sup>b</sup> mmol/L	.14	(34.34)	.13	.18	(29.84)	.17
C-reactive protein>5 mg/L	0.23	(32.22)	0.21	0.20	(27.18)	0.16
Fibrinogen<1.7 g/L	.01	(41.61)	.01	.02	(37.46)	.02
Fibrinogen>3.7 g/L	.11	(41.61)	.11	.1	(37.46)	.08
Normal blood pressure untreated	.63	(22.93)	.75	.6	(22.41)	.72
Normal blood pressure treated	.17	(22.93)	.09	.18	(22.41)	.1
High blood pressure treated	.09	(22.93)	.05	.07	(22.41)	.04
Underweight (BMI<20)	.04	(11.79)	.03	.02	(10.46)	.01
Overweight (25≤BMI<30)	.38	(11.79)	.4	.52	(10.46)	.53
Obese (30≤BMI)	.25	(11.79)	.24	.21	(10.46)	.21
Waist-Hip-Ratio> 1 <sup>a</sup> , 0.85 <sup>b</sup>	.29	(16.31)	.25	.15	(13.84)	.14
N	4,648		1,574	3,556		1,242

**Notes:**

A - Entire sample, B - Part of sample with no item non-response and no attrition

<sup>1</sup> 12-item General Health Questionnaire, values range from 0 to 12.<sup>2</sup> For different longstanding illnesses see Table 3.<sup>3</sup> Variable ranges from 0 to 16 (0 to 15 in B) for women, and 0 to 16 (0 to 12 in B) for men.<sup>4</sup> Variable ranges from 0 to 8 (0 to 5 in B) for women, and 0 to 5 (0 to 3 in B) for men.<sup>5</sup> Variable ranges from 0 to 104 (0 to 30 in B) for women, and 0 to 144 (0 to 36 in B) for men.<sup>6</sup> Reference ranges taken from Oliveira (2008)

a) Value for men, b) Value for women



Table 3: Prevalence of Different Longstanding Illnesses in Wave 0

Variable	Women		Men	
	A Mean	B Mean	A Mean	B Mean
Cancer	.03	.02	.02	.01
Endocrine/metabolic	.05	.05	.02	.02
Mental illness	.03	.02	.02	.02
Migraine/headaches	.02	.02	.01	.01
Other problem nervous system	.03	.02	.03	.02
Cataract/poor eye sight	.02	.01	.02	.02
Other eye problems	.01	.01	.02	.02
Poor hearing/deafness	.02	.01	.03	.03
Other ear complaints	.04	.01	.05	.02
Embolic	.01	.01	.02	.01
Bronchities/emphysema	.01	.01	.02	.01
Asthma	.06	.06	.05	.06
Respiratory complaints	.02	.02	.03	.03
Stomach ulcer	.03	.02	.03	.04
Other digestive complaints	.02	.02	.01	.01
Complaints of Bowel/colon	.04	.03	.02	.02
Reproductive system disorders	.01	.02	.02	.02
Arthritis	.18	.15	.12	.12
Back problems	.06	.06	.08	.08
Problems of bones/joints/muscles	.07	.06	.08	.09
Skin complaints	.01	.01	.02	.01
N	4,648	1,574	3,556	1,242

*Notes:* Information on specific longstanding illness is missing for 0.09 % of women and 0.06% of men.

A - Entire sample

B - Part of sample with no item non-response and no attrition

Table 4: Information Used in Underwriting

	Health Risk		Mortality Risk	
	Group HI <sup>a</sup>	Individual HI <sup>b</sup>	Annuities <sup>c</sup>	Life Insurance <sup>d</sup>
Age		✓	✓	✓
Sex		✓	✓	✓
Smoking Behavior		✓		✓
Medical Conditions		✓		✓
Objective Health Data		✓		✓
Use of Prescription Drugs		✓		✓
Prior Use of Medical Services		✓		
Family Health History		(✓)		✓
Occupational Class		✓		✓
Alcohol/Substance Abuse		✓		✓
Driving Information		✓		✓
Residence/Citizenship		✓		✓
Dangerous Hobbies		✓		✓
Foreign Travel		✓		✓

*Notes:*

*a)* See Cutler et al. (2008) and Anon. (1999). *b)* See Anon. (1999). *c)* See Cutler et al. (2008). *d)* See Cutler et al. (2008) and He (2009). The table displays types of information used in risk classification and calculation of premia in different insurance markets. While in the US family history is typically not used in underwriting health insurance (Anon. (1999)) it is used in the UK.

Table 5: Private Information on Mortality Risk – Women

	1	2	3	4	5	6	7
Controls							
SRH – very bad/bad	.119** (.067)	.150*** (.055)	.141*** (.053)	.071* (.044)	.079** (.045)	.079** (.044)	.055* (.028)
SRH – good	-.138*** (.021)	-.07*** (.023)	-.065*** (.023)	-.048** (.025)	-.049** (.025)	-.05** (.025)	-.03 (.015)
SRH – very good	-.141*** (.020)	-.079*** (.024)	-.073*** (.024)	-.050* (.028)	-.052* (.028)	-.054* (.028)	-.03 (.015)
Age splines		✓	✓	✓	✓	✓	✓
Smoking Behavior			✓	✓	✓	✓	✓
Medical Conditions				✓	✓	✓	✓
Family Health History					✓	✓	✓
Occupational Class						✓	✓
Objective Health Data							✓
Wald test (p-value)	.000	.000	.000	.019	.012	.009	.070
N	2,184	2,184	2,184	2,155	2,132	2,132	2,132
Pseudo R <sup>2</sup>	0.0679	0.3030	0.3266	0.3712	0.3811	0.3877	0.4256

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

Notes: Dependent variable is 1 if individual known to be dead by 2008. Reference category: SRH – fair. Average partial effects after probit estimation calculated as  $\frac{1}{N} \sum_i \Phi(x_i\hat{\beta} + \widehat{\beta_{SRH}}) - \Phi(x_i\hat{\beta})$ . Standard errors in parentheses. Estimates are weighted to correct for selection due to missing objective health data. Wald test p-values for testing the hypothesis  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ . Medical Conditions include number of prescription drugs taken. The number of observations changes between columns as observations for which some explanatory variables result in perfect prediction in the probit estimation are automatically dropped.

Table 6: Private Information on Mortality Risk – Men

	1	2	3	4	5	6	7
Controls							
SRH – very bad/bad	.246*** (.065)	.203*** (.064)	.193*** (.062)	.102** (.056)	.102** (.055)	.094** (.053)	.058 (.032)
SRH – good	-.158*** (.025)	-.097*** (.030)	-.091*** (.029)	-.068** (.031)	-.063** (.030)	-.059** (.030)	-.047** (.024)
SRH – very good	-.192*** (.022)	-.136*** (.031)	-.125*** (.030)	-.065** (.033)	-.058* (.032)	-.055* (.032)	-.033 (.018)
Age splines		✓	✓	✓	✓	✓	✓
Smoking Behavior			✓	✓	✓	✓	✓
Medical Conditions				✓	✓	✓	✓
Family Health History					✓	✓	✓
Occupational Class						✓	✓
Objective Health Data							✓
Wald test (p-value)	.000	.000	.000	.003	.006	.011	.044
N	1,766	1,766	1,766	1,766	1,766	1,766	1,766
Pseudo R <sup>2</sup>	0.1051	0.3206	0.3350	0.3961	0.4043	0.4088	0.4389

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

Notes: Dependent variable is 1 if individual known to be dead by 2008. Reference category: SRH – fair. Average partial effects after probit estimation calculated as  $\frac{1}{N} \sum_i \Phi(x_i\beta + \beta_{SRH}) - \Phi(x_i\hat{\beta})$ . Standard errors in parentheses. Estimates are weighted to correct for selection due to missing objective health data. Wald test p-values for testing the hypothesis  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ . Medical Conditions include number of prescription drugs taken.

Table 7: Private Information on Health Risk – Women

Controls	1	2	3	4	5	6	7	8
SRH – very bad/bad	.078 (.137)	.129 (.104)	.124 (.102)	-.084 (.055)	-.088 (.056)	-.085 (.055)	-.088 (.057)	-.099* (.089)
SRH – good	-.175*** (.043)	-.091** (.049)	-.091** (.050)	.010 (.033)	.010 (.033)	.009 (.032)	.008 (.033)	.011 (.013)
SRH – very good	-.243*** (.038)	-.176*** (.062)	-.177*** (.063)	-.023 (.036)	-.022 (.036)	-.019 (.036)	-.021 (.036)	-.011 (.014)
Age splines		✓	✓	✓	✓	✓	✓	✓
Smoking Behavior			✓	✓	✓	✓	✓	✓
Medical Conditions				✓	✓	✓	✓	✓
Service Use					✓	✓	✓	✓
Family Health History						✓	✓	✓
Occupational Class							✓	✓
Objective Health Data								✓
Wald test (p-value)	.000	.000	.000	.248	.238	.284	.273	.244
N	1,572	1,572	1,572	1,572	1,572	1,572	1,572	1,571
Pseudo R <sup>2</sup>	0.0519	0.1381	0.1389	0.2551	0.2557	0.2636	0.2651	0.2801

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

Notes: Dependent variable is 1 if individual diagnosed or re-diagnosed with heart disease, cancer or stroke in waves 1,2 or 3. Reference category: SRH – fair. Average partial effects after probit estimation calculated as  $\frac{1}{N} \sum_i \Phi(x_i\hat{\beta} + \beta_{SRH}) - \Phi(x_i\hat{\beta})$ . Standard errors in parentheses. Estimates are weighted to correct for selection due to missing objective health data and attrition. Wald test p-values for testing the hypothesis  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ . Medical Conditions include number of prescription drugs taken. The number of observations changes between columns as observations for which some explanatory variables result in perfect prediction in the probit estimation are automatically dropped

Table 8: Private Information on Health Risk – Men

Controls	1	2	3	4	5	6	7	8
SRH – very bad/bad	.102 (.120)	.064 (.104)	.048 (.098)	-.056 (.080)	-.046 (.074)	-.025 (.064)	-.027 (.062)	-.021 (.035)
SRH – good	-.071 (.060)	-.05 (.060)	-.049 (.055)	.008 (.043)	.008 (.041)	.011 (.038)	.010 (.037)	.019 (.030)
SRH – very good	-.123* (.060)	-.106* (.065)	-.108* (.058)	-.026 (.049)	-.023 (.047)	-.007 (.041)	-.010 (.041)	-.002 (.009)
Age splines		✓	✓	✓	✓	✓	✓	✓
Smoking Behavior			✓	✓	✓	✓	✓	✓
Medical Conditions				✓	✓	✓	✓	✓
Service Use					✓	✓	✓	✓
Family Health History						✓	✓	✓
Occupational Class							✓	✓
Objective Health Data								✓
Wald test (p-value)	.093	.170	.180	.652	.695	.893	.849	.771
N	1,240	1,240	1,240	1,240	1,240	1,240	1,240	1,239
Pseudo R <sup>2</sup>	0.0224	0.0650	0.0784	0.2217	0.2243	0.2495	0.2520	0.2746

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

Notes: Dependent variable is 1 if individual diagnosed or re-diagnosed with heart disease, cancer or stroke in waves 1,2 or 3. Reference category: SRH – fair. Average partial effects after probit estimation calculated as  $\frac{1}{N} \sum_i \Phi(x_i\hat{\beta} + \beta_{SRH}) - \Phi(x_i\hat{\beta})$ . Standard errors in parentheses. Estimates are weighted to correct for selection due to missing objective health data and attrition. Wald test p-values for testing the hypothesis  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ . Medical Conditions include number of prescription drugs taken.

Table 9: Robustness – Mortality Risk and Subjective Life Expectancy

	Women		Men	
	Annuity	Life Insurance	Annuity	Life Insurance
<u>Subjective Life Expectancy</u>				
Probability of surviving – 0 – 24%	.005 (.012)	.007 (.020)	.021 (.017)	.018 (.031)
Probability of surviving – 50 – 74%	-.023** (.013)	-.059** (.052)	-.046*** (.019)	-.038*** (.061)
Probability of surviving – 75 – 100%	-.019 (.013)	-.040 (.040)	-.044** (.021)	-.036*** (.059)
Wald test (p-value)	0.018	0.016	0.000	0.003
N	2,565	2,504	2,142	2,136
Pseudo R <sup>2</sup>	0.1408	0.2814	0.1417	0.3073
<u>Self-Rated Health</u>				
SRH – bad/fair	.019** (.009)	.023 (.023)	.074*** (.015)	.032*** (.052)
SRH – very good	-.028** (.013)	-.039 (.038)	.026 (.018)	-.011 (.021)
SRH – excellent	-.013 (.013)	-.021 (.033)	.006 (.021)	.004 (.011)
Wald test (p-value)	0.000	0.131	0.000	0.004
N	2,565	2,504	2,141	2,135
Pseudo R <sup>2</sup>	0.1622	0.2734	0.1769	0.3081

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

*Notes:* Standard errors in parentheses. Dependent variable is whether an individual is known to be dead by the year 2008. Explanatory variables are taken from ELSA wave 2. Reference category of subjective life expectancy is 25-49 % chance of survival up to certain age. SRH measured on the US scale in wave 2, good SRH, the middle category, chosen as reference. The different columns include different sets of control variables. Column Annuities includes age splines, Life Insurance includes age splines, smoking information, medical conditions, number of prescription drugs taken, family health history, and objective health data. Results are weighted to correct for missing objective health data using weights provided in ELSA.

Table 10: Robustness – Heterogeneous Effects and Follow-up Time

	Mortality Risk		Health Risk	
	Annuity	Life Insurance	Group HI	Individual HI
<u>Women younger than 65</u>				
SRH – very bad/bad	.065 (.072)	-.03 (.015)	-.043 (.065)	-.073 (.086)
SRH – good	-.063* (.032)	-.015 (.009)	-.06 (.039)	-.003 (.011)
SRH – very good	-.079** (.031)	-.001 (.011)	-.131*** (.03)	-.092** (.097)
<u>Men younger than 65</u>				
SRH – very bad/bad	.195*** (.088)	.064 (.021)	.087 (.141)	.08** (.102)
SRH – good	-.093** (.038)	-.044 (.012)	-.152** (.06)	-.006 (.01)
SRH – very good	-.124*** (.039)	-.02 (.013)	-.24*** (.043)	-.031 (.045)
<u>Women – event by 2002</u>				
SRH – very bad/bad	.096*** (.044)	.043 (.022)	.1 (.15)	-.032 (.033)
SRH – good	-.031 (.019)	-.002 (.006)	-.174*** (.042)	-.003 (.008)
SRH – very good	-.03 (.02)	.003 (.007)	-.209*** (.036)	-.004 (.01)
<u>Men – event by 2002</u>				
SRH – very bad/bad	.094** (.051)	-.003 (.01)	.127 (.115)	.004 (.013)
SRH – good	-.041* (.021)	-.013 (.008)	-.086 (.052)	-.013 (.019)
SRH – very good	-.082*** (.02)	-.027 (.013)	-.154** (.046)	-.078** (.098)

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

*Notes:* Standard errors in parentheses. Panels 1 and 2 show average marginal effects after probit estimation for individuals younger than 65. Panels 3 and 4 shorten the follow-up time to 4 years. Different columns include different sets of control variables. Column Annuities includes age splines, Life Insurance includes age splines, smoking information, medical conditions, number of prescription drugs taken, family health history, social occupational class, and objective health data, Group HI includes no controls in addition to SRH, and Individual HI includes the same variables as Life Insurance plus use of medical services. Results are weighted to correct for missing data.



## Figures

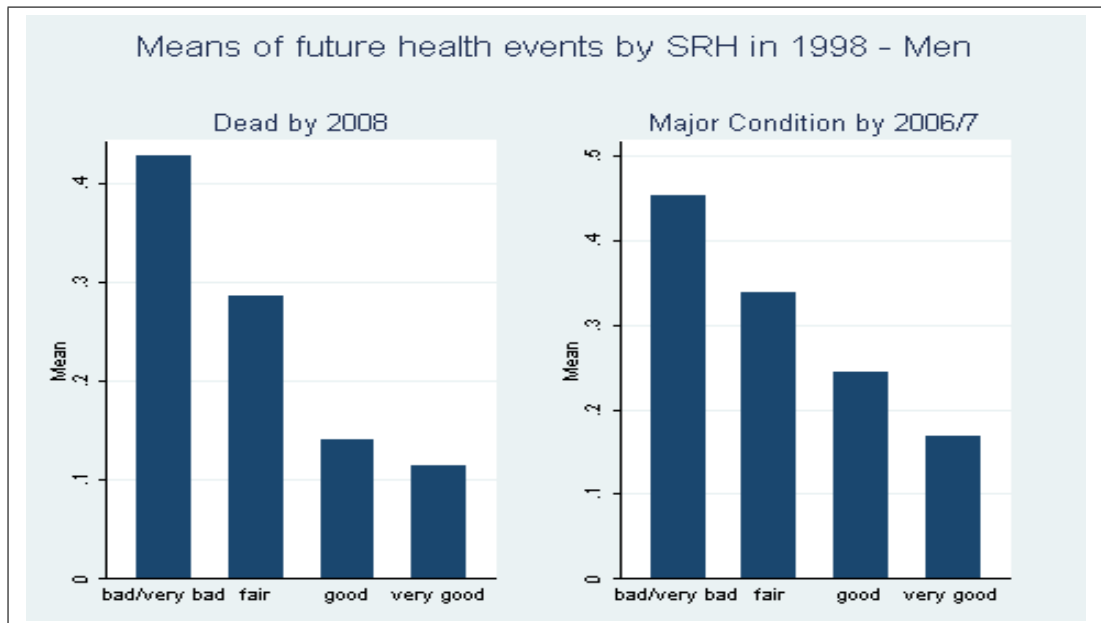


Figure 1: SRH and future health events - Men

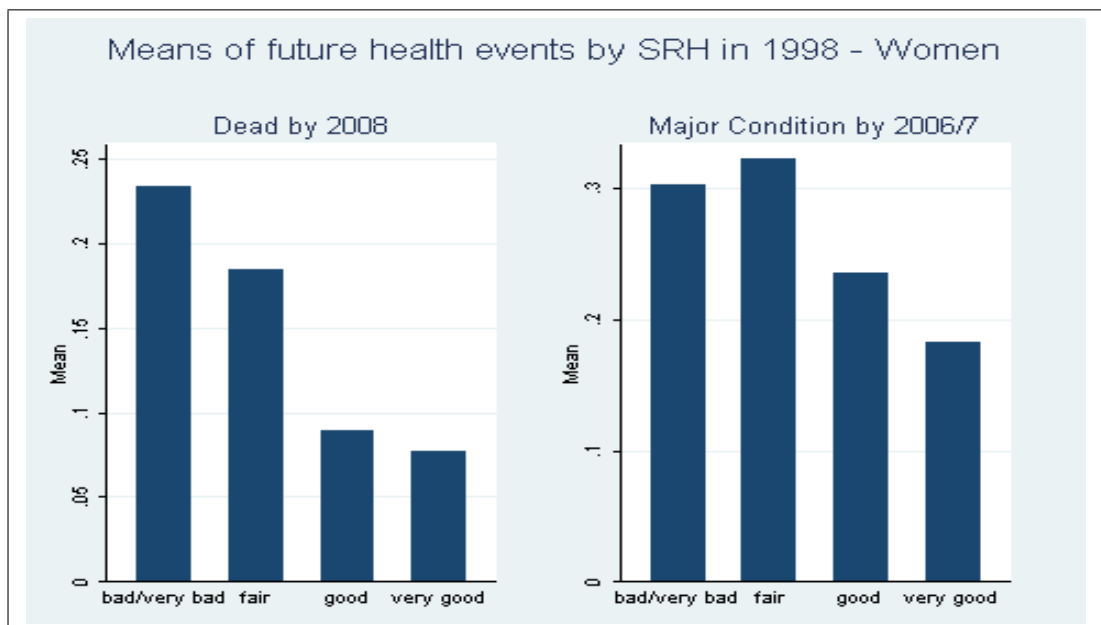


Figure 2: SRH and future health events - Women

# Appendix

## A – Inverse Probability Weighting

In order to correct for item non-response and attrition inverse probability weighting is employed. To correct for the two selection mechanisms simultaneously, I estimate the joint probability of availability of all objective health measures in wave 0, ( $h = 1$ ), and no attrition after wave 0, ( $a = 1$ ), using a bivariate probit model.

The crucial assumption for consistency with this approach is conditional independence:

$$Pr(a = 1, h = 1|y, SRH, X_{IndividualHI}, d) = Pr(a = 1, h = 1|SRH, Z, d)$$

where  $Z \subset X_{IndividualHI}$ , and  $X_{IndividualHI}$  is the largest set of underwriting controls.  $Z$  includes all variables in  $X_{IndividualHI}$  except for the objectively measured health data.  $d$  is a vector of additionally included control variables.

The conditional independence assumption would be invalidated by the existence of unobservables that influence both, selection and the outcome  $y$ . The variables included in  $d$  thus not only have to be significant predictors of attrition and item non-response but also have to be related to  $y$  in order to attenuate the worry of unobservable influences. Potential candidates for inclusion in  $d$  are health-related variables that are not used by insurance companies for underwriting and are therefore not included in the different sets of control variables  $X$ .

In my analysis,  $d$  includes information on marital status, activity status, race, the household's economic situation, survey participation behavior of other survey members in the same household, the individual's survey participation behavior in other parts of the survey in wave 0, and information on the situation during the interview.

Separate bivariate probit models are estimated for men and women. The coefficients are displayed in Table 11. Many of the variables included in  $d$  significantly affect selection. Furthermore, the two selection mechanisms are positively correlated for both genders. An individual who is more likely to stay in the survey is thus also more likely to have a nurse visit and a blood sample taken. This could reflect an unobserved liking for surveys.

Table 11: Selection Mechanisms - Bivariate Probit

	Women				Men			
	a		h		a		h	
Completed survey	.197	(.167)	.377**	(.157)	-.146	(.202)	.465***	(.180)
Partner a=1	2.36***	(.067)	.262***	(.049)	2.468***	(.071)	.235***	(.053)
Partner h=1	-.154**	(.064)	.638***	(.049)	-.112*	(.067)	.603***	(.051)
Interviewed alone	.233***	(.063)	.103*	(.055)	.599***	(.080)	.058	(.064)
HH size	-.064**	(.030)	.004	(.026)	-.035	(.035)	.009	(.027)
Rent home	-.05	(.058)	.016	(.052)	-.100	(.074)	-.131**	(.063)
Married	-.73***	(.076)	-.245***	(.067)	-.609***	(.096)	-.295***	(.081)
Widowed	-.034	(.075)	-.006	(.072)	.310***	(.113)	.081	(.106)
White	.118	(.125)	.045	(.115)	-.242	(.153)	.023	(.129)
Retired	-.018	(.068)	.016	(.059)	-.17	(.213)	-.195	(.186)
Unemployed	.623***	(.235)	.082	(.188)	-.484*	(.254)	-.462**	(.221)
Working	-.039	(.075)	.036	(.061)	-.273	(.200)	-.299*	(.176)
Self-Rated Health	✓		✓		✓		✓	
Age splines	✓		✓		✓		✓	
Smoking Behavior	✓		✓		✓		✓	
Medical Conditions	✓		✓		✓		✓	
Service Use	✓		✓		✓		✓	
Family Health History	✓		✓		✓		✓	
Occupational Class	✓		✓		✓		✓	
$\rho$	.233*** (.029)				.149*** (.035)			
N	4,648				3,556			

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

*Notes:* Standard errors in parentheses.  $a$  is equal to 1 if an individual is observed at least once after ELSA wave 0,  $h$  is equal to 1 if all objectively measured health variables are observed for an individual. Columns 1 and 2 (3 and 4) report coefficients after bivariate probit analysis for women (men).

The results in Table 11 are used to predict  $Pr(a = 1, h = 1|SRH, Z, d)$  and  $Pr(h = 1|SRH, Z, d)$  for each individual.  $\frac{1}{\hat{Pr}(a=1, h=1|SRH, Z, d)}$  is used as a weight in estimation when the dependent variables is an indicator of a diagnosis or reoccurrence of a major health condition. Whether an individual dies or not is observed irrespective of attrition, therefore  $\frac{1}{\hat{Pr}(h=1|SRH, Z, d)}$  is used as a weight in the estimations for private information on mortality risk.