HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 14/15

Private information in life insurance, annuity and health insurance markets

Amelie C. Wuppermann

August 2014

york.ac.uk/res/herc/hedgwp

Private information in life insurance, annuity and health insurance markets

Amelie C. Wuppermann*

This Version: May 2014

Abstract

Economic theory predicts that private information on risks in insurance markets leads to adverse selection. To counterbalance private information insurers collect and use information on applicants to assess their risk and calculate premiums in an underwriting process. Using data from the English Longitudinal Study of Ageing (ELSA) this paper documents that differences in the information used in underwriting across life insurance, annuity and health insurance markets attenuate private information to different extents. The results are in line with - and might help to reconcile - the mixed empirical evidence on adverse selection across these markets.

JEL-Code: D82, I13

Keywords: ELSA, private information, health-related risks, insurance, biomarkers

*University of Munich, Ludwigstr. 33, 80539 Munich, Germany. Email: amelie.wuppermann@econ.lmu.de, Tel.:+49 89 2180 6291

1 Introduction

Information asymmetries about risk types in competitive insurance markets are known to induce inefficient outcomes due to adverse selection. In particular, the seminal model by Rothschild and Stiglitz (1976) and various extensions robustly predict that if individuals have private information on their risks, higher risk individuals buy more insurance coverage than lower risk individuals (Chiappori et al., 2006). However, the empirical evidence of adverse selection is mixed (see Cohen and Siegelman, 2010, for an overview), leading to doubts on the existence of private information – or at least putting into question whether individuals use their private information in their decision to buy insurance coverage.

To counterbalance individual information, insurers collect and use information on insurance applicants to assess their risks in the process of insurance underwriting. Whether there is private information on risks that are insured in different insurance markets thus depends not only on how much individuals know about their risks but also on how much insurers know through underwriting. In this paper, I test empirically whether individuals have information on their own health risks and whether this information remains private after underwriting in different health-related insurance markets. That is, I test whether individuals have information on their risks that exceeds the information that insurers collect and use for underwriting. I then conduct an additional analysis to test whether individuals use their residual private information in the decision to purchase insurance coverage.

I focus on information on risks insured in health insurance, life insurance and annuity markets. The rationale for choosing these specific markets is that life insurance and annuity markets on the one hand and different health insurance markets on the other hand insure similar risks, but the evidence of adverse selection is mixed across these markets: there is strong evidence of adverse selection in annuity markets (Finkelstein and Poterba, 2002, 2004; McCarthy and Mitchell, 2010; Einav et al., 2010; Finkelstein and Poterba, 2014) but only limited evidence for life insurance markets (Cawley and Philipson, 1999; Hendel and Lizzeri, 2003; McCarthy and Mitchell, 2010)[1], and there is evidence of adverse selection

[1]However, He (2009) presents evidence of adverse selection in life insurance.

in group health insurance markets, such as employer-sponsored insurance in the U.S., but only little evidence for individual health insurance markets (see Cutler and Zeckhauser 2000 for an early summary of the literature).

The differences in adverse selection across life insurance and annuity markets can be explained by multiple dimensions of private information. Private information on risk type and on risk preferences constitutes a prominent example (Finkelstein and McGarry, 2006). Risk averse individuals tend to buy insurance coverage and try to prevent risk and thus may live longer. This counteracts adverse selection in life insurance but exacerbates adverse selection in annuity markets (Cutler et al., 2008). It is harder to argue that multiple dimensions of private information explain differences in adverse selection between health insurance markets that insure exactly the same risk. Cohen and Siegelman (2010) suggest that the absence of useful private information in some but not in other markets or individuals' inability or failure to act upon private information may explain difference in adverse selection across markets or market segments. In this paper I contribute to this literature by documenting that underwriting differences across health insurance markets indeed attenuate private information to different extents. Furthermore, I present evidence that individuals act upon their private information if they can. As there is typically no underwriting in group health insurance markets but strict underwriting in the individual health insurance market, my results are in line with the mixed evidence of adverse selection across these different types of markets.

Underwriting differences across markets are at least partly due to legal restrictions. U.S. federal law, for example, prohibits employers from pricing premiums for health insurance based on health-related information (GAO, 2003). Similarly, several U.S. states have introduced community rating in combination with guaranteed issue laws for individual health insurance, entitling individuals to buy insurance coverage at a price depending exclusively on abstract criteria such as sex, age or geographic region (see Lo Sasso and Lurie, 2009, for an analysis of the impact of these rules). In private individual health insurance markets in other U.S. states and countries like the U.K. and Germany health-related information is widely used in underwriting. Recent regulations in both Europe and the U.S. have,

however, limited the information that can be used for underwriting. A 2011 ruling of the European Court of Justice bans the use of gender in insurance underwriting in EU countries as of December 2012. Similarly, the new U.S. health insurance exchanges legislated by the Patient Protection and Affordable Care Act in the 2010 U.S. health reform feature limited risk rating based on age, geographic region, household size and tobacco use alone.

The analyses in this paper are based on data from the English Longitudinal Study of Ageing (ELSA), a panel study that allows to track individuals over time. The panel structure allows for observing whether individuals experience events in the future that would result in insurance claims if individuals were insured. I call this the future 'realization of risk'. In the main analysis, I interpret it as evidence for private information if a measure of an individual's self-rated health (SRH) helps to predict the future realization of risk when all information that the insurer uses in underwriting is controlled for. This approach is similar to the one of Finkelstein and McGarry (2006) who measure private information in long term care insurance using the subjective probability of entering a nursing home as a proxy for individual information on the insured risk.

The ELSA dataset contains a broad range of health measures that correspond to the information collected and used by insurers for underwriting. ELSA is one of the few longitudinal datasets that includes information on biomarkers, i.e. health data which are objectively measured and reported by a nurse. In particular, results of a blood sample analysis, blood pressure measurement, objectively measured body mass index (BMI), and waist-hip-ratio are available in ELSA. As the objectively measured data are already available in an early wave of the survey, up to 10 years of follow-up can be used for the analysis.

I measure the realization of the risk insured in life insurance and annuity markets by an indicator for whether an individual dies within 10 years after the baseline interview. The realization of the risk insured 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 in the 10-year follow up period. These conditions belong to the most costly conditions at the per capita level (Druss et al., 2002) and thus seem to be reasonable proxies for the risk insured in health insurance markets, namely high medical expenses.

My results indicate that SRH contains information on dying within or surviving the next 10 years and on being diagnosed with one or more of the costly major health conditions in the next 10 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 the biomarker data, however, the explanatory power of SRH is reduced. The explanatory power of SRH for the onset of the costly health conditions even vanishes completely. These results are robust to the choice of the proxy for individual information, to focusing on shorter time horizons and to excluding individuals with preexisting conditions from the sample. This suggests that individuals have information on the risks insured in life insurance, annuity and health insurance markets with limited underwriting, but that little of this information remains private if insurers use more extensive - in particular medical - information for underwriting.

Whether the detected private information is relevant for insurance markets in reality depends on whether it translates into actual insurance purchases. To interpret the findings I analyze whether SRH predicts insurance purchases of the different insurance products. The results of this analysis indicate that SRH predicts the purchase of annuities, but not of life insurance and private health insurance. As strict medical underwriting is employed in the latter two markets in England, these results are not surprising: Individuals might just lack private information to act upon. The results for the annuity market, however, show that SRH contains information which is correlated with actual insurance purchases. The information in SRH thus predicts actual decision making on insurance.

The paper is structured as follows: Section 2 introduces the data used for the analysis. Section 3 outlines the estimation strategy, results are shown in Section 4, Section 5 reports the results of several sensitivity analyses, Section 6 investigates whether SRH predicts insurance purchases, and the last section concludes.

2 Data

The English Longitudinal Study of Ageing (ELSA) is a rich panel dataset 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 first 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 HSE years, called ELSA wave 0, I use data from ELSA waves 1, 2, 3 and 4 collected in 2002/3, 2004/5, 2006/7 and 2008/9, respectively.[2]

While ELSA was highly influenced by and modeled on the U.S. Health and Retirement Study (HRS), its design differs from that of the HRS in one important feature: In addition to the biannual interview, every four years a nurse visit has been conducted as part of ELSA from the start.[3] From these nurse visits objectively measured blood pressure, results of a blood sample analysis, and anthropometric data are available. 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 contain 8,267 individuals from HSE 1998 (7,807 core sample members and 459 younger partners). In principle, everyone who was interviewed was eligible for the nurse visit in the HSE. Core sample members and younger partners from wave 0 are therefore included in this analysis. As the exact age is not known for individuals older than 90, I exclude these individuals from the analyses. This reduces the sample size by less than 1% to 8,205 and has no significant effects on the results. As the key variable of interest is self-reported health, one additional individual for whom this information is

[2]For a more thorough description of ELSA see Marmot et al. (2011).

[3]The HRS has also started to collect biomarker information on a subsample of its respondents in 2006. Up to now, however, a shorter follow-up is possible with the HRS than with ELSA.

not available is dropped from the analysis, reducing the sample size to 8,204.

— TABLE 1 ABOUT HERE —

Tables 1, 2, 3, and table A-1 in Appendix A display descriptive statistics for the samples used in the different analyses. The first column of each table shows the means of the variables for all available observations in the entire sample of 8,204 individuals. The second column of each of these tables shows the percent of missing observations for each variable. While demographics reported in table 1 and self-reported health information reported in table 2 are only missing for very few individuals, column 2 of table 3 indicates that some of the biomarkers are missing for almost 40% of individuals. This stems mainly from the fact that the measurement of objective health data is not compulsory in ELSA. Instead, individuals can refuse to participate in the nurse visit, and even if they agree to the nurse visit they can refuse to have a blood sample taken. Overall, all objective health measures in wave 0 are available only for about 50% of the sample. I call this the sample with 'biomarker participation'. As biomarker information is important to imitate medical screening in insurance underwriting, I limit my analyses to individuals in this sample.

— TABLE 2 ABOUT HERE —

Descriptive statistics for the sample with biomarker participation are displayed in column 3 of tables 1, 2, and 3. Comparing the means in column 3 to the means of the entire sample in column 1 indicates that the selection into biomarker participation might be systematic. The sample with biomarker participation is on average younger, in a higher social-occupational class and less likely to smoke. 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 selection 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. To correct for this, I develop and employ an inverse probability weighting strategy that is discussed in detail in appendix B.

When analyzing a new diagnosis or recurrence of a major health condition in the future the sample size shrinks further due to attrition. While information on whether individuals die is collected regardless of attrition by linking the data to information from the U.K. Department of Work and Pensions and to information contained in the National Health Service Central Register held by the Office of National Statistics, information on future diagnosis or recurrence of the major diseases is derived from individuals' answers to questions in the 2002, 2004, 2006 and 2008 waves of ELSA. The information is thus only available for individuals who appear again in ELSA after 1998. Overall, of the 3,950 individuals for whom the objective health data are available in 1998 only 71% are observed at least once in 2002, 2004, 2006 or 2008. For the analyses that focus on the future diagnosis or recurrence of major health conditions, there are thus two selection mechanisms that reduce the sample size: On the one hand, individuals have to participate in the nurse visit and have to have a blood sample taken. On the other hand, they have to stay in ELSA in the later waves, i.e. there has to be 'no attrition'.

— TABLE 3 ABOUT HERE —

The fourth columns of tables 1, 2, and 3 display means of the variables for the sample that participates in the biomarker collection and does not completely drop out of ELSA after wave 0. Compared to the sample that only requires biomarker participation this sample is even younger and healthier. In the analysis that focuses on realization of health risks and is thus subject to the two selection mechanisms, biomarker participation and no attrition, the weights for inverse probability weighting are derived based on a bivariate probit model that accounts for the two selection mechanisms simultaneously (see appendix B for details).

The SRH measure in ELSA is of particular importance for my analysis as it is employed as the main proxy for individuals' information on their health risks. Individuals in ELSA wave 0 were asked "How is your health in general? Would you say it was very good, good, fair, bad, or very bad?" Column 1 of table 2 displays mean responses for the different SRH categories. As there are only few individuals who rate their health as bad or very bad, I

group these two categories into one.

Figure 1 presents a first glance at the relationship between SRH and the future health events. There are clear graded relationships of SRH and both outcome measures. While 40% of individuals with bad or very bad health in 1998 have died by 2008, only about 10% of individuals who rate their health as very good in 1998 have died. Similarly, almost 50% of individuals with bad or very bad health in 1998 have experienced an onset or a recurrence of heart disease, cancer or stroke by ELSA wave 4, while the fraction is only 20% among individuals in very good health in 1998.

The relationships between the future health events and SRH suggest that individuals have information on the risks insured in life insurance, annuity and health insurance markets. In the next section, I explore whether this information remains when the information that insurers collect and use in underwriting across the different markets is controlled for and thus whether individuals have information on the insured risks that is additional to what the insurers know.

3 Estimation Strategy

The existence of private information on the risks insured in health insurance, life insurance and annuity markets is investigated by regressing indicator variables for the occurrence of future health events on categories of SRH and different control variables in a baseline year. I include different sets of control variables to capture the information used for underwriting by insurance companies in the different insurance markets. The future health events are meant to capture the future realization of risks. The realization of the risk insured in life insurance and annuity markets is captured by an indicator for whether an individual is dead 10 years after the baseline interview. The realization of the risk insured 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 in the 10-year follow-up period.

The information on the future health events that is contained in SRH is interpreted as

evidence for private information. There are two reasons for the choice of SRH as a proxy for individual information: First, it has been shown that SRH significantly predicts future health events, like death and the onset of specific health conditions.[4] If individuals who rate their own health as worse are truly in worse health and have a higher probability to need health care and to die sooner, SRH captures useful information on the risks insured in the insurance markets that I focus on. 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 co-morbidities or family health history which can potentially 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-related risks that is only contained in SRH necessarily remains private. SRH is a relatively coarse measure, however, and might not capture all relevant information on health and mortality risk. In a robustness analysis, I therefore analyze how well the underwriting controls are able to predict the realization of the insured risk and thus the general scope for private information in the different markets.

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

$$\begin{aligned} y_i^{j*} &= \beta_{0a}^j + \beta_{1a}^j SRH_{1i} + \beta_{2a}^j SRH_{2i} + \beta_{3a}^j SRH_{3i} + \boldsymbol{\beta_{4a}^j X_{ai}} + \epsilon_{ai}^j \qquad (1) \\ y_i^j &= \mathbb{I}(y_i^{j*} > 0) \end{aligned}$$

where SRH_k, $k \in \{1, 2, 3\}$, represent dummy variables for the three SRH categories: very bad/bad, good and very good, with fair SRH as reference category. $\boldsymbol{X_{ai}}$ is a $K \times 1$ vector of variables that represents the information which insurance companies in insurance market a, $a \in$ (Annuities, Life Insurance, Group Health Insurance, Private Individual

[4]For overviews on studies analyzing the relationship between SRH and subsequent death see Idler and Benyamini (1997) and DeSalvo et al. (2006). Banks et al. (2007) provide evidence for relationships between SRH and future diagnoses of diseases.

Health Insurance), collect and use for underwriting. $\boldsymbol{\beta_{4a}}$ is the $1 \times K$ vector of coefficients associated with $\boldsymbol{X_{ai}}$.

ϵ_{ai}^{j} captures unobservables influences on the latent future health event j with underwriting controls a. Under the assumptions that $\epsilon_{ai}^{j} \sim N(0,1)$ and that the correlation between ϵ_{a}^{D} and ϵ_{a}^{M}, ρ_{DM}, is equal to 0 for set of underwriting controls a, I estimate the vector of all coefficients, $\boldsymbol{\beta_a} = (\beta_{0a}, \beta_{1a}, \beta_{2a}, \beta_{3a}, \boldsymbol{\beta_{4a}})$, for each set of underwriting controls and each of the health events independently using single equation probit models.[5] The coefficients $\beta_{1a}^{j}, \beta_{2a}^{j}$, and β_{3a}^{j} measure whether SRH helps to explain the future realization of risks when underwriting controls from insurance market a are included in the model. If the coefficients are (jointly) equal to zero, there is no information in SRH that is additional to the information contained in $\boldsymbol{X_a}$ and thus no evidence for private information in market a - as captured by SRH.

— TABLE 4 ABOUT HERE —

Table 4 displays the variables that are typically collected in the application process and used for risk classification and calculation of premiums in different insurance markets in the U.K. and in the U.S. As Finkelstein and Poterba (2014) observe, insurance companies may have additional information about applicants that they do not use for underwriting. These "unused observables" do not mitigate the scope for adverse selection and are thus not included as controls in $\boldsymbol{X_a}$.

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. In the U.S. employer-sponsored health insurance markets, U.S. federal law forbids individual underwriting (GAO, 2003). In the U.K., employer-sponsored private health insurance also refrains from individual underwriting (Mossialos and Thomson, 2009). In annuity markets, underwriting has historically been based on the individual's age and sex. This information

[5]The assumption of independence between the error terms can be relaxed and a bivariate equation probit model estimated. For this analysis, however, the assumption of independent error terms seems appropriate as the interest lies in 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 thus included in the vector of control variables $\boldsymbol{X}_{Annuities}$. In recent years, however, underwriting information in annuity markets has been extended. In both the U.S. and the U.K. medically underwritten annuities are available to sick individuals or smokers. In the U.K. this process started at the end of the 20th century. Furthermore, at least since 2007 information on postcodes has been used for underwriting U.K. annuities (Finkelstein and Poterba, 2014). I include the additional information when testing for the use of private information in annuity purchases in the U.K. in section 6.

Considerably more information is used for underwriting life and individual health insurance. In addition to age and sex, information on lifestyle is typically used. Age, sex and information on an individual's smoking history are thus included in the vector of control variables. Furthermore, medical information of individuals and sometimes also their families can be used for underwriting. The insurers in the U.K. and some U.S. states can even require insurance applicants to undergo medical examinations. They may thus have clinical information on blood values and other objectively measured information. To capture the medical information I include self-reported conditions, linear splines of objectively measured BMI, waist-to-hip ratio, and blood values in $\boldsymbol{X}_{Life}$ and $\boldsymbol{X}_{IndividualHI}$. I further use parents' cause of death as a proxy for familial medical history.

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 occupation, I include dummies for occupational social class as represented in table 1. The other conditions are unfortunately not well captured in the ELSA data.

4 Results

Average marginal effects of the SRH categories after estimation of equation (1) are reported in tables 5 and 6. The first table displays average marginal effects of the SRH categories on the probability of dying within the next 10 years. Table 6 displays average marginal effects of SRH on the probability of a new diagnosis or recurrence of one of the major

health conditions, heart disease, cancer or stroke, in the next 10 years. Both tables show results that are weighted to correct for missing data.[6]

— TABLE 5 ABOUT HERE —

The results presented in column 2 in table 5 are of specific interest. In addition to SRH the estimated model includes the annuity underwriting controls age and sex. All three SRH coefficients are significantly different from 0. Conditional on age and sex, individuals in very good SRH in 1998 are 9 percentage points less likely to be dead in 2008 and individuals in good health 7 percentage points less likely to be dead in 2008 than individuals who rate their health as fair. For individuals who rate their health as bad or very bad in 1998, in contrast, the probability of having died by 2008 is 17 percentage points higher than for individuals with fair SRH. When interpreting conditional information in SRH as private information these results show evidence of a scope for adverse selection in annuity markets.

From left to right in table 5 more control variables are added. Even when smoking status and medical conditions are included as underwriting controls in column 4, there is additional information in SRH that predicts death. The recent inclusion of smoking information and preexisting conditions in underwriting annuities thus likely did not eliminate private information in this market. The last column of table 5 reports average marginal effects of the SRH categories when age, sex, smoking information, medical conditions, family health history, social occupational class, and objective health data are included as controls. The p-value of the Wald test for joint significance indicates that the three SRH coefficients are still jointly significantly different from 0 at the 5 % 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 SRH coefficients. Thorough underwriting, and in particular medical underwriting that includes blood tests and other objectively measured health data, is thus able to reduce private information about mortality risks.

— TABLE 6 ABOUT HERE —

[6]The estimation of the weights is described in appendix B. Unweighted results are qualitatively similar to the weighted ones. They are displayed in tables C-1 and C-2 in Appendix C.

Table 6 shows similar results for the diagnosis or re-diagnosis of a major health condition within 10 years after the baseline interview. The results presented in column 1 indicate that with no controls added in addition to SRH, the coefficient of the three SRH categories are jointly significantly different from zero. When there is no individual underwriting, as in the case of group health insurance, there is thus private information and therefore scope for adverse selection.

The inclusion of additional controls from left to right in table 6 results in a reduction in the information in SRH. The inclusion of self-reported medical conditions and the number of prescription drugs taken leaves no additional explanatory power in SRH for the diagnosis or re-diagnosis of a major health condition as the results presented in column 4 indicate. Taking information in SRH on a future diagnosis of a major condition as private information, the results indicate that medical underwriting is a crucial determinant of the amount of private information left in health insurance markets.

Overall, the results present evidence that medical underwriting, and in particular underwriting including medical examinations significantly reduces private information as captured by SRH. While I find evidence that private information on mortality risks remains even with stringent underwriting, I find no evidence of private information on health risks when stringent underwriting is employed. There is evidence, however, for private information on health risks when no or scarce information is used in underwriting.

5 Robustness Analyses

In this section, I present three sensitivity analyses. First, I reestimate equation (1) for relatively healthy individuals to investigate whether private information among the sick drives the main results. Second, I shorten the time horizon within which the realization of risk can occur to 4 years to investigate whether individuals have more information on risks in a shorter term. Third, I analyze how well the underwriting controls help to predict the future realizations of risk to gauge the scope for private information independent of a proxy for individual information.

— TABLE 7 ABOUT HERE —

Table 7 displays average marginal effects of the SRH categories after estimating equation (1) with samples of relatively healthy individuals, i.e. with individuals without specific preexisting health conditions. Hendren (2013) suggests that sicker individuals hold more information on their health risks than healthier ones. The evidence on private information with limited underwriting presented in the last section might thus be driven by sick individuals. However, if sick individuals are deterred from applying for these insurance products, for example by exclusions of, or waiting periods for, preexisting conditions, private information among these individuals might not be relevant for the market in reality.[7] By excluding individuals with preexisting conditions I thus investigate whether there is private information also among healthier individuals. Furthermore, in markets with strict underwriting individuals with preexisting conditions might not be able to purchase insurance at all because their applications are rejected by the insurer (see Hendren, 2013). Following the classification by Hendren (2013), I thus exclude all individual who might be rejected during the application process and analyze private information among individuals who will likely be able to buy insurance even in the case of strict underwriting.

Results presented in the upper panel of table 7 are based on individuals who have never had a diagnosis of a heart condition or a stroke and who don't currently have cancer. The results in the lower panel are based on a sample who in addition do not suffer from diabetes, chronic lung disease, or hypertension. Hendren (2013) suggests also to exclude individuals with a former diagnosis of cancer. This information, however, is unfortunately not available in ELSA wave 0. Columns (1) and (2) present estimates for mortality risk, columns (3) and (4) for health risks. The first column for each event includes a limited set of control variables in addition to SRH. The column labeled Annuities corresponds to column (2) in table 5, while the column labeled Group HI corresponds to column (1) in table 6. The second column for each analyzed risk includes the most comprehensive sets of underwriting controls, $\boldsymbol{X}_{Life}$ and $\boldsymbol{X}_{IndividualHI}$. They correspond to the last columns of

[7] In the U.S. group health insurance, however, waiting periods for preexisting conditions are limited by federal law through the Health Insurance Portability and Accountability Act of 1996.

tables 5 and 6, respectively.

The results presented in table 7 show that individuals who do not currently have cancer and have not been diagnosed with any of the two other conditions before the baseline year and rate their health as very good are about 11 percentage points less likely to be diagnosed with one of the conditions in the next 10 years than individuals in the same sample who rate their health as fair. The evidence of private information in SRH with limited underwriting is thus not driven by individuals with cancer or a prior diagnosis of a stroke or heart disease. When also excluding individuals with diabetes, chronic lung disease or hypertension, the information in the SRH categories is no longer jointly significantly different from 0. Similar to Hendren (2013) I thus find evidence that healthier individuals have less private information than sick ones.

The results for mortality risk presented in table 7 indicate that SRH has no significant explanatory power for dying within the next 10 years with stringent underwriting for individuals whose application will not be rejected by the insurer due to a preexisting condition. The information in SRH on mortality risk that is detected in the main analysis thus seems to be concentrated among individuals with preexisting conditions. This finding is in line with Hendren (2013)'s finding that individuals with preexisting conditions have private information on the insured risks, while healthier individuals do not. As applications by individuals with preexisting conditions are often rejected only individuals without additional private information are able to purchase insurance which supports the idea that stringent underwriting limits private information in life insurance markets.

The results for the second sensitivity analysis are presented in table 8. This table reports average marginal effects of the three SRH categories when the dependent variables capture the realization of risk within the next 4 years. The dependent variable in columns 1 and 2 is an indicator for whether an individual is dead by ELSA wave 1. In columns 3 and 4 the dependent variable is an indicator for whether an individual reports a diagnosis or recurrence of a major health condition in wave 1.

Similar to the results of the main analysis, there is significant information in SRH on dying within the next 4 years and on being diagnosed or re-diagnosed with a major health

condition within the next 4 years when only limited sets of underwriting controls are included. This information is eliminated when the most comprehensive set of underwriting controls, $\boldsymbol{X}_{Life}$ for mortality risk and $\boldsymbol{X}_{IndividualHI}$ for health risk, is included in the estimation. There is thus no evidence that SRH contains more information that is additional to the full set of underwriting controls in the shorter than in the longer term.

— TABLE 8 ABOUT HERE —

SRH is only a proxy for individual information on health and mortality risks. Individuals could have information on the insured risks that is neither included in SRH nor in the verifiable underwriting controls. In order to gauge the scope for private information independent of a proxy for individual information, I analyze how well the underwriting controls allow to predict the realizations of the risks. For this, I estimate probit models specified in equation 2 that only include the different sets of underwriting controls and omit proxies for individual information:

$$\begin{aligned} y_i^{j*} &= \gamma_{0a}^j + \gamma_{1a}^j \boldsymbol{X_{ai}} + \eta_{ai}^j \\ y_i^j &= \mathbb{I}(y_i^{j*} > 0). \end{aligned} \tag{2}$$

$\boldsymbol{X_{ai}}$ represents the vectors of the respective most comprehensive sets of underwriting controls for both dependent variables, y_i^M and y_i^D, and
$eta_a^j \sim N(0,1)$. The better these comprehensive sets of underwriting controls help to predict the future realization of the insured risk, the better the insurer will be able to discriminate between risk types based on the collected information. The less scope thus remains for private information.

In order to analyze how well models explain the variation in binary dependent variables, the percent of correctly predicted observations is often reported. This approach involves predicting the conditional probability of observing an outcome of one, $\widehat{Pr}(y_i^j = 1|\boldsymbol{X_{ai}})$, for each individual and choosing some cutoff above which the researcher assumes that the predicted probability corresponds to predicting a 1. One can then calculate how many of the actual ones and how many of the actual zeros are correctly predicted by the model. It

is not clear, however, what the optimal choice of cutoff is (see Wooldridge, 2010, p. 573f., for a discussion).

In order to avoid the choice of a specific cutoff, I display the distributions of the predicted probabilities by the actual values of the outcome variable. Figure 2 contains two empirical cumulative distribution functions (cdfs) of the conditional probability of being dead 10 years after the baseline, $\widehat{Pr}(y_i^D = 1|\boldsymbol{X}_{Life})$, one for individuals who are still alive in 2008, cdf_0, and one for individuals who are dead by 2008, cdf_1. Analogously, figure 3 displays the empirical cdfs of $\widehat{Pr}(y_i^M = 1|\boldsymbol{X}_{IndividualHI})$ by actual onset or recurrence of a major health condition.

These figures allow to determine the percent of correctly predicted ones and zeros for each possible cutoff, $t \in [0, 1]$. As for the construction of the percent of correctly predicted outcomes with one specific cutoff, I assume that $\widehat{Pr}(y_i = 1|\boldsymbol{X}_{ai}) > t$ corresponds to predicting a one and $\widehat{Pr}(y_i = 1|\boldsymbol{X}_{ai}) \leq t$ corresponds to predicting a zero. For each cutoff t, the value of the cdf of the predicted probabilities at t corresponds the share of *predicted zeros*, and one minus the value of the cdf at t corresponds to the share of *predicted ones*. As the cdfs in figures 2 and 3 are displayed separately for individuals with an actual outcome of one, cdf_1, and with an actual outcome of zero, cdf_0, the figures display the share of *correctly predicted ones* and the share of *correctly predicted zeros* at each cutoff t. The share of correctly predicted ones corresponds to one minus the value of cdf_1 at t, $1 - cdf_1(t)$, while the share of correctly predicted zeros corresponds to the value of cdf_0 at t, $cdf_0(t)$.[8]

— FIGURE 2 ABOUT HERE —

— FIGURE 3 ABOUT HERE —

In figure 2 with a cutoff at 20%, for example, 80% of individuals who are still alive in 2008 have a predicted probability that is equal to or lower than the cutoff, $cut_0(0.2) \approx 0.8$,

[8]The idea of determining how well a model predicts the actual outcome for all possible cutoffs is also employed in receiver operating characteristic (ROC) analysis. This technique is frequently used in medical and epidemiological research, for example in evaluations of diagnostic tests. See Zou et al. (2007) and references therein for an introduction to ROC analysis.

and 80% of individuals who are dead by 2008 have a predicted probability that is higher than the cutoff, $1 - cut_1(0.2) \approx 0.8$. If an insurer thus set a cutoff at 20% and treated every individual with a predicted probability below this cutoff as not likely to die within the next 10 years and every individual with a predicted probability above this cutoff as likely to die, it would correctly classify 80% of individuals. Figure 3 displays a similar pattern for the model that predicts the onset or recurrence of a major health conditions based on the full set of underwriting controls.

Figures 2 and 3 suggest that it seems to be possible to discriminate rather accurately between individuals for whom the risks materialize in the future and individuals for whom the risks do not materialize based on a simple probit model using the full set of underwriting controls. There thus does not seem to be a lot of room for private information in the different insurance markets when stringent underwriting is employed. As this analysis does not use a proxy for individual information, the result holds independent of the choice of a proxy for individual information.

Overall, the analyses presented in this section strengthen the intuition that stringent underwriting - in combination with rejections of insurance applications based on medical history - helps to eliminate private information in the markets for life insurance and individual health insurance. In particular, there is no evidence for private information on the risk of dying in the next 10 years among healthy individuals who will likely be able to obtain life insurance coverage when stringent underwriting is employed. The private information detected in the main analysis seems to be limited to individuals whose applications will be rejected by life insurers. Furthermore, the main results are robust to changing the time horizon and independent of the proxy used for individual information.

6 SRH and the Purchase of Insurance

Whether the detected private information translates into adverse selection in the actual insurance markets depends on the question whether individuals act upon this information in their decision to buy insurance. In this section I analyze whether SRH helps to predict

insurance purchases in the ELSA data. I focus on buying insurance products rather than correlation of insurance status and SRH in the entire population to avoid the selection problem outlined by He (2009). Furthermore, in the case of health insurance SRH measured after the decision to purchase insurance coverage might be affected by insurance status and thus not capture the relevant individual information.

The ideal data for this analysis would contain information on individuals' intention or wish to buy insurance in addition to information on the actual purchases. In particular in markets with stringent underwriting and possible rejections of applications, the two can differ because individuals who demand insurance might not be able to obtain it. Individuals with low SRH might thus not be able to buy health or life insurance even though they want to. There might thus be no correlation between SRH and actual insurance purchase even though SRH captures information that individuals would act upon if they could.

The ELSA data, however, do not contain information on the intention to purchase insurance. They do not even contain direct information on actual insurance purchases. Starting from ELSA wave 1, however, the dataset contains information on whether individuals receive annuity income, whether they hold life insurance, and whether they hold individual private health insurance, that is private health insurance which is not payed for by their employer. Based on this information, I construct indicators for whether individuals who do not receive annuity income or do not hold the respective insurance, buy a new annuity product, life insurance or individual private health insurance in the future. Among the individuals who do not hold the respective insurance I code those as buyers who report holding it in a later wave. While the insurance information is available as of ELSA wave 1, wave 2 is the first wave with both, biomarker information and insurance information. Using ELSA wave 2 as baseline allows for an analysis of the relationship of SRH and insurance purchases conditional on medical underwriting information.

— TABLE 9 ABOUT HERE —

Table 9 reports average marginal effects of the three categories of SRH after probit estimation of the purchase of the three different insurance products. Columns (1) and (2)

refer to buying an annuity in waves 2 or 3, columns (3) and (4) to buying life insurance, and columns (5) and (6) to buying individual private health insurance. In the first column of each type of insurance the relation between SRH and insurance purchase without additional controls is investigated. In the second column the largest available set of underwriting controls for the respective market is included.[9]

The results in table 9 show that the purchase of a new annuity is significantly related to SRH, while purchasing life insurance and individual private health insurance is not significantly predicted by SRH. These results are independent of whether the underwriting controls are included in addition to SRH. The results without additional controls indicate that those individuals who rate their health as worse are not more likely to buy life or health insurance, while those who rate their health as better are more likely to buy annuities. To the extent that insurance companies can charge different prices based on objective observable health measures, it is not important whether sicker individuals buy more or less insurance but whether among the individuals that the insurance treats as equal - based on their observables - sicker individuals are more likely to buy insurance. I thus include the underwriting control in this analysis. As insurers in the U.K. annuity market are using additional information to age and sex since the early 2000s, such as smoking status and health information, I include additional variables in the underwriting controls. In addition, insurers have started using information on an applicant's residence in the period that I am looking at (Finkelstein and Poterba, 2014). As there is no address information in my data, I cannot include this type of information. However, the results presented in table 9 show that even when neglecting postcode information individuals who rate their health as bad or very bad purchase annuities with a significantly lower probability. As insurers have an incentive to sell more annuities to sicker individuals, it is unlikely that the relationship between SRH and the probability to buy an annuity only reflects information that is captured in postcode data and is thus known to the insurer. Thus, individuals seem

[9]Unfortunately, not all control variables that were used in the main analysis are available in ELSA wave 2. In particular, there is neither information on the use of medical services, such as hospital stays or GP visits, nor on the cause of parents' death. Underwriting controls can thus not be mimicked as well with wave 2 as with wave 0.

to use the information on their health in the decision to buy annuities.

There are two main conclusions that can be drawn from these results. First, SRH does contain information that predicts the purchase of annuities and thus SRH is related to actual decisions on insurance products. Second, the presented results support the empirical evidence of no adverse selection in life insurance and private health insurance markets in the literature, as individuals who newly buy life insurance or private health insurance do not seem to be less healthy than individuals who do not buy insurance.

7 Discussion and Conclusion

The empirical literature has found mixed evidence on adverse selection in different insurance markets. Even between markets that insure similar risks, namely the markets for life insurance and annuity on the one hand and different health insurance markets on the other hand, the evidence on adverse selection varies. In this paper, I document that underwriting differences across these markets lead to different amounts of private information on the insured risks. Furthermore, in markets without stringent underwriting, individuals seem to use their private information in the decision to purchase insurance products. As the markets with stringent underwriting (life insurance and individual private health insurance) show no evidence of adverse selection, while markets with less stringent or even no underwriting (annuities and group health insurance) appear to be adversely selected, these results suggest that differences in underwriting may – at least partly – help to explain differences in adverse selection between markets.

In the interpretation of these results, it should be borne in mind that they rely on several assumptions. First, I assume that controlling for underwriting information captures what insurers know about individual risks. If insurers used the information more or less effectively, my estimates of private information would not measure the actual private information in the insurance market. However, at least for another health-related insurance market, the market of long term care insurance, the analysis of Finkelstein and McGarry (2006) indicates that controlling for underwriting information delivers similar results as

controlling for the insurers' actuarial prediction based on this information.

Second, I assume that individual information on risks is captured by SRH. Of course, this is only a crude proxy and individuals could know more about their risks than reflected in SRH. My analysis might thus not detect all private information. However, I show that the underwriting controls in markets with stringent underwriting are able to predict the actual outcomes rather well. Insurance companies should thus be able to predict risks well based on the underwriting information leaving only little room for private information and adverse selection in markets with stringent underwriting.

My results further suggest that stringent underwriting can eliminate private information in the markets that insure mortality and health risks. Employing stringent underwriting could thus also reduce and possibly eliminate adverse selection. Recent research indicates that adverse selection reduces welfare in the market for annuities (Einav et al., 2010) and limited risk-rating of premiums reduces welfare in health insurance markets (Bundorf et al., 2012). A question that my analysis cannot answer, however, is how stringent underwriting would affect overall welfare in these markets. On the one hand, stringent underwriting might increase welfare by eliminating welfare losses resulting from adverse selection. On the other hand, it might at the same time reduce welfare in other dimensions. For example, in the case of short-term insurance contracts where individuals have to buy insurance coverage repeatedly, repeated stringent underwriting would introduce an additional risk, namely the risk of future 'risk reclassification'. If there is no insurance for this additional risk, a "welfare loss from incomplete insurance contracts" will arise (Cutler and Zeckhauser, 2000, p. 627). The welfare effects of stringent underwriting thus likely depend on the specific insurance setting.

References

Banks, J., Crossley, T., and Goshev, S. (2007). Looking for Private Information in Self-assessed Health. HEDG Working Paper, University of York 9.

British Medical Association (2008). Medical information and insurance: Joint guidelines from the British Medical Association and the Association of British Insurers.

Bundorf, M., Levin, J., and Mahoney, N. (2012). Pricing and Welfare in Health Plan Choice. *American Economic Review*, 102(7):3214–3248.

Cawley, J. and Philipson, T. (1999). An empirical examination of information barriers to trade in insurance. *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., and McGarry, K. (2008). Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance. NBER Working Paper 13746.

Cutler, D. and Zeckhauser, R. (2000). The Anatomy of Health Insurance. In Culyer, A. and Newhouse, J., editors, *Handbook of Health Economics*, volume I, pages 563–643. Elsevier.

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.

Druss, B., Marcus, S., Olfson, M., and Pincus, H. (2002). The most expensive medical conditions in America. *Health Affairs*, 21(4):105–111.

Einav, L., Finkelstein, A., and Schrimpf, P. (2010). Optimal mandates and the welfare cost of asymmetric information: evidence from the UK annuity market. *Econometrica*, 78(3):1031–1092.

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. (2014). Testing for Adverse Selection using "Unused Observables" in Insurance Markets: Evidence from the U.K. Annuity Market. *The Journal of Risk and Insurance*, 9999(9999):1–26.

GAO (2003). Private Health Insurance: Federal and State Requirements Affecting Coverage Offered by Small Business. United States General Accounting Office.

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.

Hendren, N. (2013). Private information and insurance rejections. *Econometrica*, 81(5):1713–1762.

Idler, E. and Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior*, 38(1):21–37.

Lo Sasso, A. and Lurie, I. (2009). Community rating and the market for private non-group health insurance. *Journal of Public Economics*, 93(1-2):264–279.

Marmot, M., Banks, J., Blundell, R., Erens, B., Lessof, C., Nazroo, J., and Huppert, F. (2011). English Longitudinal Study of Ageing: Wave 0 (1998, 1999 and 2001) and Waves 1-4 (2002-2009). 15th Edition. Colchester, Essex: UK Data Archive [distributor] SN: 5050, http://dx.doi.org/10.5255/UKDA-SN-5050-1.

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.

Risk Classification in Individually Purchased Voluntary Medical Expense Insurance (1999). The American Academy of Actuaries Issue Paper.

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.

Wooldridge, J. (2010). *Econometric analysis of cross section and panel data*. The MIT press, 2 edition.

Zou, K., O' Malley, A., and Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115(5):654–657.

Tables

Acknowledgments

This research has profited from comments by seminar and conference participants at the Universities of Darmstadt, Munich, and Tilburg, at the ifo Institute, at the ASHecon Conference in Cornell, at the ECHE in Helsinki, at the ESWC in Shanghai and the EEA in Glasgow. In addition, I received insightful comments from Sebastian Bauhoff, Luc Bissonnette, Florian Heiss, Hendrik Jürges, Helmut Farbmacher, Daniel McFadden, Erik Meijer, Nicolas Sauter, Beatrice Scheubel and Joachim Winter. Financial support from the Deutsche Forschungsgemeinschaft through GRK 801 is gratefully acknowledged. The data used in this analysis were made available through the U.K. Data Archive. ELSA was developed by a team of researchers based at the National Centre for Social Research, University College London and the Institute for Fiscal Studies. The data were collected by the National Centre for Social Research. The funding is provided by the National Institute of Aging in the United States, and a consortium of U.K. government departments coordinated by the Office for National Statistics. The developers and funders of ELSA and the Archive do not bear any responsibility for the analyses or interpretations presented here.

Table 1: Demographics – ELSA Wave 0

Variable	(1) Overall	(2) % missing	(3) Bio Part	(4) Bio Part & No Attr
Demographics				
Male	0.43	0.00	0.45	0.44
Age < 50	0.18	0.00	0.21	0.21
Age 50-59	0.3	0.00	0.32	0.34
Age 60-69	0.25	0.00	0.25	0.27
Age 70-79	0.19	0.00	0.16	0.14
Age 80-89	0.08	0.00	0.06	0.04
Occupational Class				
Professional	0.06	0.09	0.07	0.08
Managerial - technical	0.3	0.09	0.31	0.32
Skilled - non manual	0.13	0.09	0.12	0.12
Skilled - manual	0.29	0.09	0.29	0.29
Semi-skilled manual	0.15	0.09	0.14	0.13
Unskilled manual	0.06	0.09	0.05	0.05
Other social class	0.02	0.09	0.01	0.01
Smoking Behavior, Reference: Never				
Ever Smoker	0.64	0.11	0.63	0.61
Current Smoker	0.22	0.11	0.19	0.18
N	8204		3950	2815

Notes:
Column 1 - Means in entire sample
Column 2 - Percent missing observations within entire sample
Column 3 - Means in part of sample with biomarker participation
Column 4 - Means in part of sample with biomarker participation and no attrition

Table 2: Health Measures – ELSA Wave 0

Variable	(1) Overall	(2) % missing	(3) Bio Part	(4) Bio Part & No Attr
Self-Rated Health				
Very bad/bad	0.1	0.00	0.07	0.06
Fair	0.24	0.00	0.21	0.2
Good	0.38	0.00	0.41	0.41
Very good	0.28	0.00	0.32	0.33
Medical Conditions				
Hypertension	0.3	0.07	0.24	0.23
Diabetes	0.05	0.01	0.04	0.03
Stroke	0.04	0.01	0.02	0.02
Heart Attack	0.05	0.02	0.03	0.02
Angina	0.08	0.01	0.05	0.04
Heart Murmur	0.04	0.01	0.03	0.03
Irregular Heart Rhythm	0.07	0.00	0.05	0.05
Other Heart Problems	0.02	0.01	0.01	0.01
GHQ12[1]	1.45	6.24	1.23	1.18
No Longstanding Illness (LI)[2]	0.44	0.05	0.49	0.5
Non-limiting LI	0.21	0.05	0.22	0.22
Limiting LI	0.36	0.05	0.29	0.28
Use of Medical Services				
# Prescription drugs taken	1.75	13.71	1.28	1.22
# of GP visits last 2 weeks	0.22	0.02	0.19	0.18
# Hospital nights last year	1.08	0.06	0.59	0.48
Family Health History				
Father dead	0.87	2.41	0.85	0.85
Mother dead	0.74	1.68	0.7	0.69
At least one parent died of				
Hypertension	0.01	1.52	0.01	0.01
Angina	0.03	1.52	0.03	0.03
Heart Attack	0.27	1.52	0.26	0.27
Other Heart Problem	0.14	1.52	0.14	0.14
Stroke	0.08	1.52	0.09	0.09
Diabetes	0.02	1.52	0.02	0.02
N	8204		3950	2815

Notes: Column 1 - Means in entire sample, Column 2 - percent missing observations in entire sample, Column 3 - Means in part of sample with biomarker participation, Column 4 - Means in part of sample with biomarker participation and no attrition. [1]12-item General Health Questionnaire, values range from 0 to 12. [2]For different longstanding illnesses see table A-1 in Appendix A.

Table 3: Biomarkers – ELSA Wave 0

Variable	(1) Entire Sample	(2) % missing	(3) Bio Part	(4) Bio Part & No Attr
Blood analysis[1]				
Haemoglobin<13[a)], 11.5[b)] g/dL	0.06	30.35	0.06	0.05
Haemoglobin>18[a)], 16.5[b)] g/dL	0.002	30.35	0.002	0.001
Ferritin< 25[a)], 20[b)] μg/L	0.11	31.52	0.11	0.11
Ferritin>400[a)], 200[b)] μg/L	0.02	31.52	0.02	0.02
Total cholesterol>5 mmol/L	0.77	31.86	0.78	0.78
HDL cholesterol<1[a)], 1.2[b)] mmol/L	0.16	32.06	0.15	0.14
C-reactive protein>5 mg/L	0.21	29.92	0.2	0.19
Fibrinogen<1.7 g/L	0.02	39.81	0.02	0.02
Fibrinogen>3.7 g/L	0.11	39.81	0.1	0.09
Blood pressure (BP)				
Normal blood pressure untreated	0.62	22.71	0.71	0.73
Normal blood pressure treated	0.17	22.71	0.1	0.09
High blood pressure treated	0.08	22.71	0.05	0.05
Body Mass Index (BMI)				
Underweight (BMI<20)	0.03	11.21	0.03	0.02
Overweight (25≤BMI<30)	0.44	11.21	0.45	0.46
Obese (30≤BMI)	0.24	11.21	0.23	0.23
Waist-Hip-Ratio> 1[a)], 0.85[b)]	0.23	15.24	0.2	0.2
N	8204		3950	2815

Notes: Column 1 - Means in entire sample, Column 2 - percent missing observations in entire sample, Column 3 - Means in part of sample with biomarker participation, Column 4 - Means in part of sample with biomarker participation and no attrition. [1]Reference ranges taken from Oliveira (2008), [a)]Value for men, [b)]Value for women.

Table 4: Information Used in Underwriting in U.S. and U.K. Insurance Markets

	Health Risk		Mortality Risk	
	Group HI[a)]	Individual HI[b)]	Annuities[c)]	Life Insurance[d)]
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		✓		✓
Postal code			(✓)	
Dangerous Hobbies		✓		✓
Foreign Travel		✓		✓

Notes:
a) See GAO (2003) and "Risk Classification" (1999) for the U.S. and Mossialos and Thomson (2009) for the U.K. *b*) See for the U.S. "Risk Classification" (1999) and and for the U.K. the British Medical Association (2008) and the Association of British Insurers at `https://www.abi.org.uk/~/media/Files/Documents/Publications/Public/Migrated/Health/ABI%20consumer%20guide%20on%20buying%20private%20medical%20insurance.ashx`, accessed May 9, 2014. *c*) See Cutler et al. (2008) for the U.S. and Finkelstein and Poterba (2014) for the U.K. *d*) See Cutler et al. (2008) and He (2009) for the U.S. and the British Medical Association (2008) for the U.K. The table displays types of information used in risk classification and calculation of premiums in different insurance markets. While in the U.S. family history is typically not used in underwriting health insurance ("Risk Classification" 1999), it maybe used in the U.K. In addition, the U.K. annuity market in the 1990s only used age and gender for insurance underwriting. In the beginning of the 21th century, however, an enhanced annuity market emerged that offered special annuities to smokers and sick individuals. Furthermore, insurers started using information on applicants address as of 2007 the latest (Finkelstein and Poterba, 2014).

Table 5: Private Information on Mortality Risk

Controls	(1)	(2)	(3)	(4)	(5)	(6)	(7)
SRH – very bad/bad	0.189***	0.172***	0.16***	0.071***	0.07***	0.064**	0.052**
	(0.047)	(0.035)	(0.033)	(0.029)	(0.029)	(0.029)	(0.027)
SRH – good	-0.145***	-0.071***	-0.065***	-0.042**	-0.041**	-0.04**	-0.03*
	(0.022)	(0.016)	(0.016)	(0.017)	(0.016)	(0.017)	(0.016)
SRH – very good	-0.166***	-0.091***	-0.082***	-0.043**	-0.042**	-0.041**	-0.028
	(0.022)	(0.017)	(0.016)	(0.018)	(0.018)	(0.018)	(0.018)
Age splines		✓	✓	✓	✓	✓	✓
Sex		✓	✓	✓	✓	✓	✓
Smoking Behavior			✓	✓	✓	✓	✓
Medical Conditions				✓	✓	✓	✓
Family Health History					✓	✓	✓
Occupational Class						✓	✓
Objective Health Data							✓
Wald test (p-value)	0.000	0.000	0.000	0.001	0.001	0.001	0.018
Pseudo R^2	0.074	0.316	0.333	0.371	0.375	0.377	0.399
Ln(L)	-3815.003	-2820.997	-2749.019	-2591.514	-2576.124	-2566.56	-2477.1
N	3950	3950	3950	3950	3950	3950	3950

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

Notes: Dependent variable is 1 if individual is dead by 2008. Reference category SRH: fair. Average marginal effects after probit estimation calculated as $\frac{1}{N}\sum_i \left[\Phi(\widehat{\beta_0}+\widehat{\beta_k}+\widehat{\beta_4}X_i)-\Phi(\widehat{\beta_0}+\widehat{\beta_4}X_i)\right]$ for $k \in \{1,2,3\}$. Standard errors of the marginal effects calculated using the delta method 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 6: Private Information on Health Risk

Controls	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SRH – very bad/bad	0.174	0.173*	0.156*	-0.0002	-0.0002	-0.003	-0.002	-0.026
	(0.107)	(0.094)	(0.089)	(0.057)	(0.056)	(0.053)	(0.053)	(0.045)
SRH – good	-0.12***	-0.072*	-0.075*	0.025	0.027	0.028	0.029	0.031
	(0.043)	(0.04)	(0.04)	(0.03)	(0.03)	(0.029)	(0.03)	(0.028)
SRH – very good	-0.182***	-0.142***	-0.147***	0.0004	0.003	0.004	0.005	0.006
	(0.044)	(0.043)	(0.042)	(0.034)	(0.034)	(0.034)	(0.035)	(0.033)
Age splines		✓	✓	✓	✓	✓	✓	✓
Sex		✓	✓	✓	✓	✓	✓	✓
Smoking Behavior			✓	✓	✓	✓	✓	✓
Medical Conditions				✓	✓	✓	✓	✓
Service Use					✓	✓	✓	✓
Family Health History						✓	✓	✓
Occupational Class							✓	✓
Objective Health Data								✓
Wald test (p-value)	0.000	0.000	0.000	0.733	0.716	0.662	0.673	0.478
Pseudo R^2	0.045	0.097	0.103	0.212	0.216	0.224	0.225	0.241
Ln(L)	-5376.24	-5086.148	-5049.905	-4439.014	-4417.741	-4369.008	-4365.457	-4276.382
N	2815	2815	2815	2815	2815	2815	2815	2815

* 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, 3 or 4. Reference category SRH: fair. Average marginal effects after probit estimation calculated as $\frac{1}{N}\sum_i \left[\Phi(\widehat{\beta_0}+\widehat{\beta_k}+\widehat{\beta_4}X_i)-\Phi(\widehat{\beta_0}+\widehat{\beta_4}X_i)\right]$ for $k \in \{1,2,3\}$. Standard errors of marginal effects calculated with the delta method 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 7: Robustness – No preexisting conditions

	Mortality Risk		**Health Risk**	
	Annuities	Life Insurance	Group HI	Individual HI
	(1)	(2)	(3)	(4)
Healthy I				
SRH – very bad/bad	0.113***	0.038	0.12	0.05
	(0.043)	(0.031)	(0.148)	(0.061)
SRH – good	-0.063***	-0.036**	-0.078*	0.004
	(0.017)	(0.016)	(0.045)	(0.032)
SRH – very good	-0.075 ***	-0.035**	-0.114**	-0.024
	(0.017)	(0.017)	(0.047)	(0.035)
Wald test (p-value)	0.000	0.026	0.032	0.555
Pseudo R^2	0.262	0.343	0.019	0.149
Ln(L)	-1927.987	-1716.61	-3817.931	-3313.413
N	3334	3334	2423	2423
Healthy II				
SRH – very bad/bad	0.138***	0.05	-0.011	-0.009
	(0.057)	(0.039)	(0.081)	(0.065)
SRH – good	-0.054***	-0.025	-0.036	-0.013
	(0.019)	(0.017)	(0.046)	(0.036)
SRH – very good	-0.064***	-0.018	-0.077*	-0.049
	(0.019)	(0.019)	(0.045)	(0.037)
Wald test (p-value)	0.004	0.128	0.222	0.444
Pseudo R^2	0.287	0.37	0.004	0.138
Ln(L)	-1262.883	-1114.72	-2488.985	-2155.077
N	2570	2570	1879	1879

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

Notes: Healthy I refers to the sample of individuals with no prior diagnosis of stroke or heart disease and who don't currently have cancer. Healthy II refers to individuals with no prior diagnosis of stroke, heart disease, and who don't have cancer, diabetes, chronic lung disease or hypertension. Average marginal effects after probit estimation and their standard errors in parentheses. Dependent variable indicator for whether an individual is dead by the year 2008 for mortality risk and indicator for whether there was an onset or recurrence of heart disease, cancer and/or stroke in waves 1, 2, 3 or 4 for health risk. Different columns include different sets of control variables. Column Annuities corresponds to column (2), Life Insurance to column (7) in table 5. Column Group HI corresponds to column (1) and Individual HI to column (8) of table 6. Results are weighted to correct for missing data.

Table 8: Robustness – 4-Year Time Horizon

	Mortality Risk		**Health Risk**	
	Annuities	Life Insurance	Group HI	Individual HI
	(1)	(2)	(3)	(4)
Event by 2002/3				
SRH – very bad/bad	0.093***	0.017	0.208*	-0.002
	(0.029)	(0.019)	(0.121)	(0.039)
SRH – good	-0.024**	-0.009	-0.13 ***	0.007
	(0.012)	(0.012)	(0.041)	(0.025)
SRH – very good	-0.04 ***	-0.015	-0.181***	-0.027
	(0.012)	(0.013)	(0.042)	(0.029)
Wald test (p-value)	0.000	0.423	0.000	0.508
Pseudo R^2	0.232	0.343	0.079	0.308
Ln(L)	-1771.864	-1515.102	-4317.545	-3240.728
N	3950	3950	2780	2780

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

Notes: Average marginal effects after probit estimation and their standard errors in parentheses. Dependent variable indicator for whether an individual is dead by 2002/3 for mortality risk, and indicator for whether there was an onset or recurrence of heart disease, cancer and/or stroke by 2002/3 for health risk. Different columns include different sets of control variables. Column Annuities corresponds to column (2), Life Insurance to column (7) in table 5. Column Group HI corresponds to column (1) and Individual HI to column (8) of table 6. Results are weighted to correct for missing data.

Table 9: SRH and Insurance Purchases

	New Annuity		New Life Insurance		New private H.I.	
	(1)	(2)	(3)	(4)	(5)	(6)
SRH – very bad/bad	-0.009**	-0.009***	0.002	-0.009	-0.001	-0.004
	(0.004)	(0.003)	(0.018)	(0.017)	(0.009)	(0.09)
SRH – good	0.001	0.002	0.033**	0.031*	0.013	0.012
	(0.005)	(0.005)	(0.016)	(0.016)	(0.009)	(0.01)
SRH – very good	0.014*	0.015*	0.022	0.018	0.02*	0.02
	(0.008)	(0.007)	(0.008)	(0.021)	(0.013)	(0.017)
Age splines		✓		✓		✓
Sex		✓		✓		✓
Smoking Behavior		✓		✓		✓
Medical Conditions		✓		✓		✓
Family Health History				✓		✓
Objective Health Data				✓		✓
Wald test (p-value)	0.003	0.004	0.164	0.146	0.178	0.254
Pseudo R^2	0.025	0.106	0.003	0.078	0.004	0.054
Ln(L)	-263.629	-241.519	–803.1898	-742.647	-615.761	-584.755
N	3954	3954	2451	2451	3594	3594

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

Notes: Analysis based on wave 2 of ELSA and follow-up. Sample of individuals who do not have respective insurance in wave 2. Dependent variable is 1 if individual obtains insurance in waves 3 or 4. Reference category SRH: fair. Average marginal effects after probit estimation. Standard errors of marginal effects calculated with the delta method in parentheses. Estimates are weighted to correct for selection due to missing objective health data using weights provided by ELSA. Wald test p-values for testing the hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

Figures

Figure 1: SRH and Future Health Events

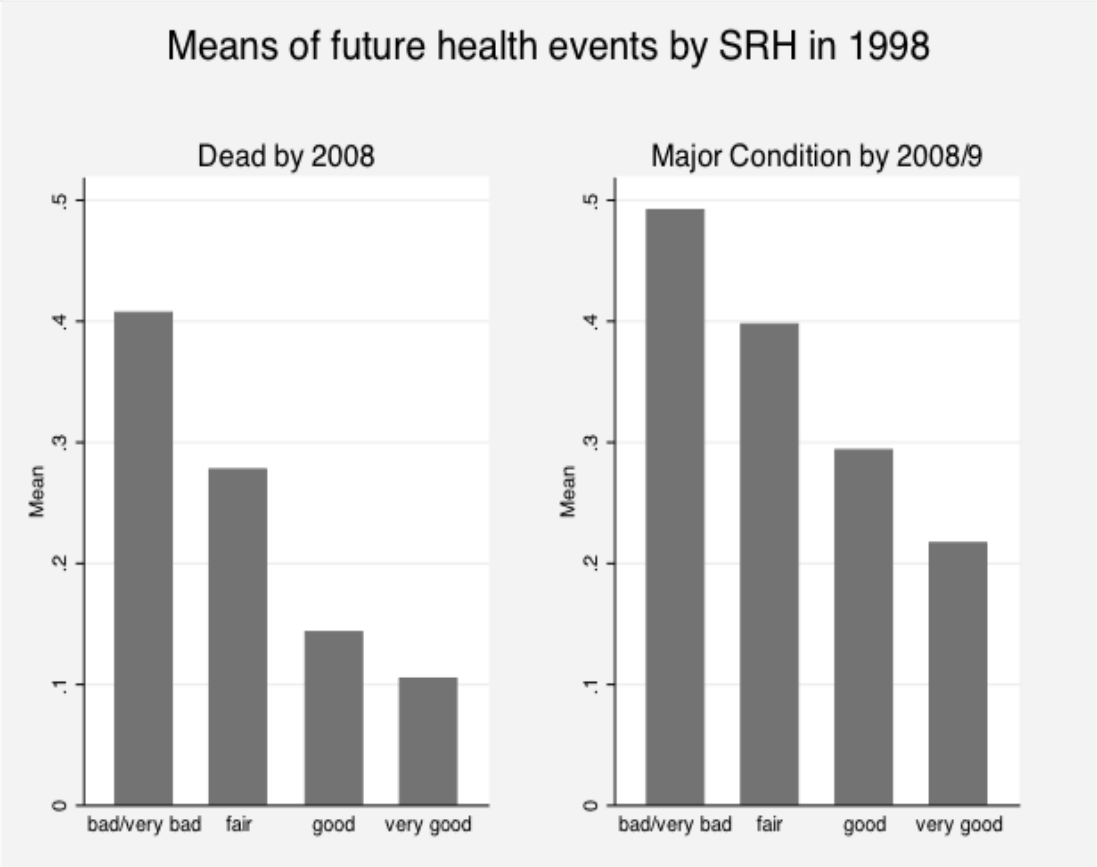


Figure 2: CDF of Predicted Probability of Death by Actual Outcome

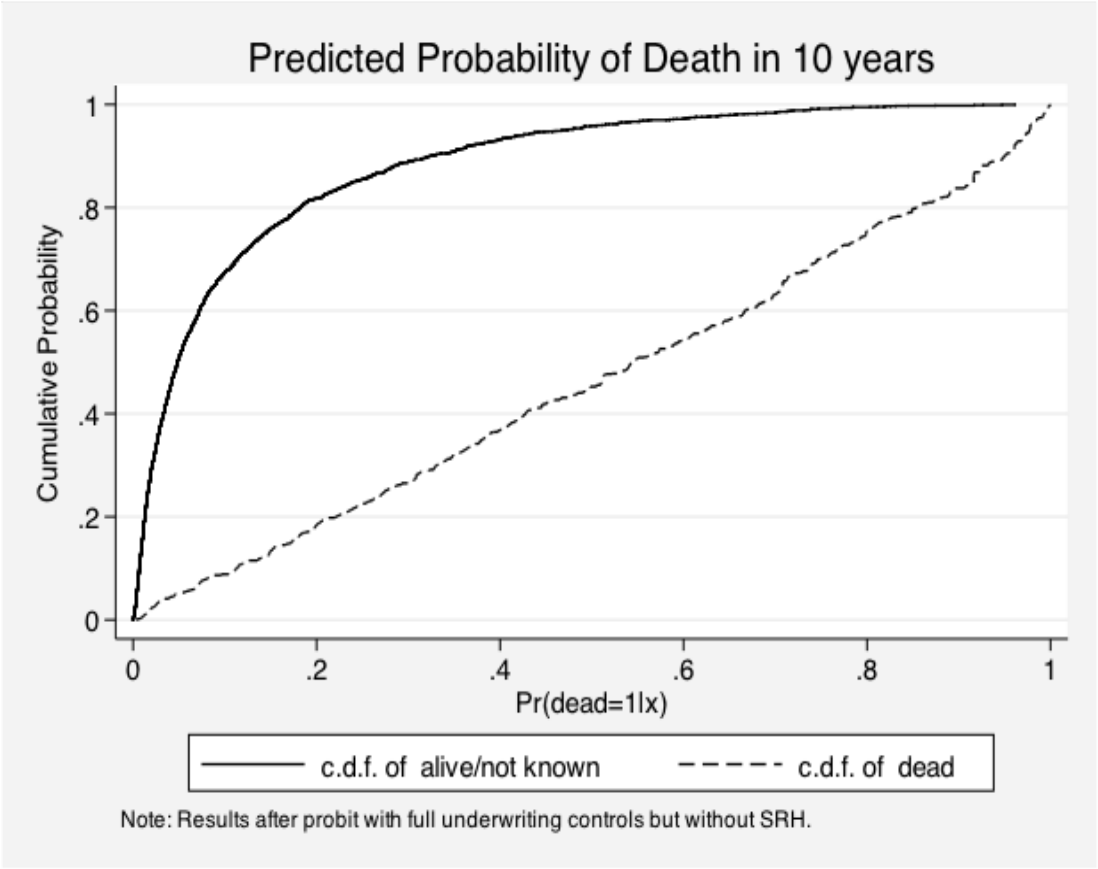
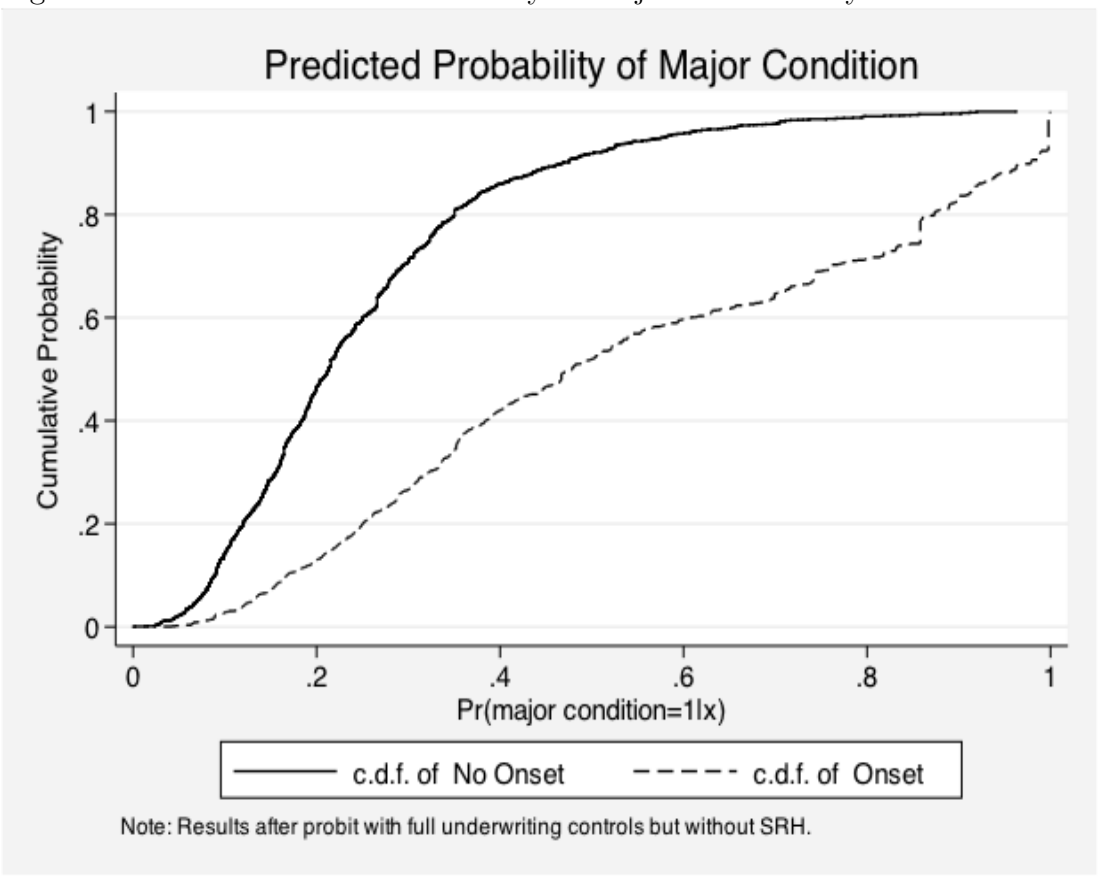


Figure 3: CDF of Predicted Probability of Major Condition by Actual Outcome



Appendix A: Longstanding Illnesses

Table A-1: Prevalence of Different Longstanding Illnesses – Wave 0

Variable	(1) Entire Sample	(2) Bio Part	(3) Bio Part & No Attr.
Cancer	0.02	0.02	0.02
Endocrine/metabolic disorders	0.04	0.03	0.03
Mental illness	0.03	0.02	0.02
Migraine/headaches	0.01	0.01	0.01
Other problem nervous system	0.03	0.02	0.02
Cataract/poor eye sight	0.02	0.02	0.01
Other eye problems	0.02	0.02	0.02
Poor hearing/deafness	0.03	0.02	0.02
Other ear complaints	0.04	0.02	0.02
Complaints of blood vessels	0.02	0.01	0.01
Bronchitis/emphysema	0.02	0.02	0.01
Asthma	0.06	0.06	0.06
Respiratory complaints	0.03	0.02	0.02
Stomach ulcer	0.03	0.03	0.03
Other digestive complaints	0.02	0.02	0.02
Complaints of bowel/colon	0.03	0.03	0.03
Reproductive system disorders	0.01	0.02	0.02
Arthritis	0.16	0.14	0.14
Back problems	0.07	0.07	0.07
Problems of bones/joints/muscles	0.08	0.07	0.07
Skin complaints	0.01	0.01	0.01
N	8198	3949	2814

Notes: Column 1 - Means in entire sample
Column 2 - Means in sample with biomarker participation
Column 3 - Means in sample with biomarker participation and no attrition

Appendix B: Inverse Probability Weighting

In order to correct for the selection into biomarker participation and no attrition I use inverse probability weighting. To correct for the two selection mechanisms simultaneously, I estimate the joint probability of biomarker participation 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

$$Pr(a = 1, h = 1|y, SRH, \boldsymbol{X}_{IndividualHI}, \boldsymbol{d}) = Pr(a = 1, h = 1|SRH, \boldsymbol{Z}, \boldsymbol{d})$$

where $\boldsymbol{Z} \subset \boldsymbol{X}_{IndividualHI}$, and $\boldsymbol{X}_{IndividualHI}$ is the most comprehensive set of underwriting controls. Z includes all variables in $\boldsymbol{X}_{IndividualHI}$ except for the objectively measured health data. $\boldsymbol{d}$ 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 $\boldsymbol{d}$ thus not only have to be significant predictors of attrition and biomarker participation but also have to be related to y in order to attenuate the worry of unobservable influences. Potential candidates for inclusion in $\boldsymbol{d}$ 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 $\boldsymbol{X}$.

In my analysis, $\boldsymbol{d}$ 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. The means of these variables are displayed in table B-1. Comparing the means displayed in column 1 with the ones in columns 3 and 4 indicates that these variables indeed seem to be related to the selection mechanisms.

The coefficients of the bivariate probit model are displayed in table B-2. Many of the variables included in $\boldsymbol{d}$ significantly affect selection. Furthermore, the two selection mechanisms are positively correlated. An individual who is more likely to stay in the survey is also more likely to have a nurse visit and a blood sample taken. This could reflect an unobserved liking for surveys.

The results in table B-2 are used to predict $Pr(a = 1, h = 1|SRH, \boldsymbol{Z}, \boldsymbol{d})$ and $Pr(h = 1|SRH, \boldsymbol{Z}, \boldsymbol{d})$ for each individual. The inverse of the predicted probability of biomarker participation and no attrition, $(\hat{Pr}(a = 1, h = 1|SRH, \boldsymbol{Z}, \boldsymbol{d}))^{-1}$, is used as a weight in the

Table B-1: Additional Controls for Inverse Probability Weighting – ELSA Wave 0

Variable	(1) Overall	(2) % missing	(3) Bio Part	(4) Bio Part & No Attr
Marital Status, Reference: Single				
Married	0.69	0.04	0.72	0.74
Widowed	0.16	0.04	0.12	0.11
Activity Status, Reference: Inactive				
Retired	0.39	0.11	0.35	0.34
Unemployed	0.02	0.11	0.02	0.02
On sick leave	0.05	0.11	0.04	0.04
Working	0.41	0.11	0.48	0.49
Race				
White	0.97	0.16	0.97	0.98
Survey Participation				
Completed other parts of survey	0.96	0.00	0.98	0.98
Partner without attrition	0.46	0.00	0.54	0.72
Partner with biomarker participation	0.36	0.00	0.48	0.51
Interviewed alone	0.34	0.16	0.31	0.29
Household Situation				
HH size	2.23	0.00	2.33	2.35
Rent home	0.21	0.00	0.17	0.15
N	8204		3950	2815

Notes:
Column 1 - Means in entire sample
Column 2 - Percent missing observations within entire sample
Column 3 - Means in part of sample with biomarker participation
Column 4 - Means in part of sample with biomarker participation and no attrition

estimations for health risks. Whether an individual dies or not is observed irrespective of attrition, therefore $(\hat{Pr}(h = 1|SRH, \boldsymbol{Z}, \boldsymbol{d}))^{-1}$ is used as a weight in the estimations of mortality risk.

Table B-2: Selection Mechanisms – Bivariate Probit

	No Attrition		Biomarker Participation	
Completed other parts of survey	0.114	(0.126)	0.419***	(0.117)
Partner without attrition	2.374***	(0.047)	0.236***	(0.036)
Partner with biomarker participation	-0.143***	(0.046)	0.604***	(0.035)
Interviewed alone	0.363***	(0.048)	0.076*	(0.041)
HH size	-0.050**	(0.022)	0.004	(0.018)
Rent Home	-0.054	(0.045)	-0.043	(0.040)
Married	-0.661***	(0.058)	-0.249***	(0.051)
Widowed	0.051	(0.061)	-0.006	(0.058)
White	-0.016	(0.095)	0.041	(0.085)
Retired	0.005	(0.062)	0.027	(0.053)
Unemployed	0.024	(0.142)	-0.148	(0.121)
On sick leave	0.162	(0.099)	-0.136	(0.087)
Working	-0.061	(0.065)	-0.022	(0.054)
SRH	✓		✓	
Age splines	✓		✓	
Smoking behavior	✓		✓	
Medical conditions	✓		✓	
Service use	✓		✓	
Family history	✓		✓	
Occupational class	✓		✓	
ρ		0.201***	(0.022)	
N		8204		

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

Notes: Coefficients after bivariate probit analysis displayed. Standard errors in parentheses. ρ captures the correlation in the unobservables between the two selection mechanisms.

Appendix C: Robustness to weighting

Tables C-1 and C-2 display marginal effects and their standard errors after unweighted probit estimation of equation (1). Except for the differences in weighting these tables are directly comparable to tables 5 and 6 in the main analysis. For both outcomes, death and the diagnosis or re-diagnosis of a major health condition, the results are qualitatively unchanged by the weighting procedure: Without further controls, SRH significantly helps to predict whether individuals die within the next 10 years and whether individuals are diagnosed or re-diagnosed with a major health condition within the next 10 years. Even with the full set of underwriting controls that are used in life insurance markets, SRH contains significant information for predicting death. Including the full set of underwriting controls for the diagnosis of the major health conditions, however, reduces the information contained in SRH to insignificant levels.

Table C-1: Private Information on Mortality Risk - No Weighting

Controls	(1)	(2)	(3)	(4)	(5)	(6)	(7)
SRH – very bad/bad	0.095***	0.095***	0.087***	0.037	0.034	0.03	0.025
	(0.032)	(0.026)	(0.025)	(0.025)	(0.024)	(0.024)	(0.023)
SRH – good	-0.116***	-0.066***	-0.06 ***	-0.039***	-0.039***	-0.038***	-0.031**
	(0.017)	(0.014)	(0.013)	(0.014)	(0.014)	(0.014)	(0.013)
SRH – very good	-0.133***	-0.083***	-0.076***	-0.043***	-0.042***	-0.041***	-0.03 **
	(0.017)	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)
Age splines		✓	✓	✓	✓	✓	✓
Sex		✓	✓	✓	✓	✓	✓
Smoking Behavior			✓	✓	✓	✓	✓
Medical Conditions				✓	✓	✓	✓
Family Health History					✓	✓	✓
Occupational Class						✓	✓
Objective Health Data							✓
Wald test (p-value)	0.000	0.000	0.000	0.001	0.002	0.004	0.032
Pseudo R^2	0.042	0.264	0.278	0.31	0.312	0.315	0.335
Ln(L)	-1563.617	-1200.965	-1178.006	-1126.78	-1122.121	-1117.797	-1085.561
N	3950	3950	3950	3950	3950	3950	3950

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

Notes: Dependent variable is 1 if individual is dead by 2008. Reference category SRH: fair. Average marginal effects after probit estimation calculated as $\frac{1}{N}\sum_i \left[\Phi(\widehat{\beta_0}+\widehat{\beta_k}+\widehat{\beta_4}X_i)-\Phi(\widehat{\beta_0}+\widehat{\beta_4}X_i)\right]$ for $k \in \{1,2,3\}$. Standard errors of the marginal effects calculated using the delta method in parentheses. 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 C-2: Private Information on Health Risk - No Weighting

Controls	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SRH – very bad/bad	0.05	0.056	0.057	-0.021	-0.028	-0.03	-0.027	-0.038
	(0.043)	(0.042)	(0.041)	(0.039)	(0.038)	(0.038)	(0.038)	(0.038)
SRH – good	-0.09 ***	-0.064***	-0.064***	-0.003	-0.002	-0.002	-0.004	-0.003
	(0.024)	(0.023)	(0.023)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
SRH – very good	-0.153***	-0.124***	-0.123***	-0.036	-0.033	-0.032	-0.035	-0.03
	(0.024)	(0.023)	(0.023)	(0.026)	(0.026)	(0.026)	(0.027)	(0.027)
Age splines		✓	✓	✓	✓	✓	✓	✓
Sex		✓	✓	✓	✓	✓	✓	✓
Smoking Behavior			✓	✓	✓	✓	✓	✓
Medical Conditions				✓	✓	✓	✓	✓
Service Use					✓	✓	✓	✓
Family Health History						✓	✓	✓
Occupational Class							✓	✓
Objective Health Data								✓
Wald test (p-value)	0.000	0.000	0.000	0.303	0.328	0.348	0.326	0.353
Pseudo R^2	0.017	0.07	0.071	0.118	0.12	0.124	0.126	0.131
Ln(L)	-1623.387	-1535.37	-1534.63	-1456.683	-1453.023	-1446.395	-1444.338	-1434.711
N	2815	2815	2815	2815	2815	2815	2815	2815

* 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, 3 or 4. Reference category SRH: fair. Average marginal effects after probit estimation calculated as $\frac{1}{N}\sum_i \left[\Phi(\widehat{\beta_0}+\widehat{\beta_k}+\widehat{\beta_4}X_i)-\Phi(\widehat{\beta_0}+\widehat{\beta_4}X_i)\right]$ for $k \in \{1,2,3\}$. Standard errors of marginal effects calculated with the delta method in parentheses. 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.