Insurance Choice and Private Information: Does Knowing Lead to Acting?

Jannis Stöckel*†‡, Bram Wouterse‡ and Pieter Bakx‡

Monday 9th June, 2025

Abstract

Individuals' perceptions of their own health might deviate from what can be observed based on objective health indicators, because of biased beliefs or private information. Whether individuals act based on their private beliefs matters. Yet, it is often hard to distinguish between what individuals think about their health and how they act upon it. We analyze insurance choice quality by studying health insurance deductible choices of Dutch adults. For this, we link the Dutch Health Monitor, a health survey with more than 1 million adult participants between 2012 and 2020, with administrative data on past and future health outcomes. Individuals' beliefs about their health are predictive for health care costs, even when controlling for extensive information on prior health care use and self-reported lifestyle behaviors. This private information is used by individuals when choosing their deductible. Yet, private information is not sufficient: many people, especially older healthier individuals do not make deductible choices in line with their private health information.

Keywords: Health insurance, private information, health perceptions, choice quality.
JEL Classification: H00, I13, G52, D81, D90

Acknowledgments: We gratefully acknowledge comments to earlier versions of this paper provided by participants at the Erasmus Initiative *"Smarter Choices for Better Health"* Conference, the LSE Department of Health Policy seminar, the lowlands Health Economists' Study Group 2024, the 2^{nd} Dutch Health Econometrics Workshop, the CINCH Health Economics Seminar, and the XV^{th} Workshop of the dggö Committee on Allocation and Distribution. This project received funding from the *"Smarter Choices for Better Health"* Seed Grant scheme. We further acknowledge usage of the Public Health Monitor (2012, 2016, and 2020) of the Community Health Services, Statistics Netherlands, and the National Institute for Public Health and the Environment. All remaining errors are our own.

*Corresponding author: Jannis Stöckel (j.stockel@lse.ac.uk)
†Department of Health Policy, LSE Health, London School of Economics
‡Erasmus School of Health Policy & Management, Erasmus University Rotterdam.

I Introduction

Do people have relevant private information and use this information when making economic decisions? These two related questions determine the welfare effects of private beliefs in economic contexts, including insurance choice. We study both questions in a setting allowing us to establish whether there is a correlation between having private information and the choices made but also to assesses the effects of private information on individual welfare by considering choice-quality. To study this we utilize large-scale survey data, administrative health records, and information on insurance choices.

Individuals may think differently about their own health than what would be predicted based on health information observable by insurers. Private information allows individuals in principle to chose insurance coverage adequate to their individual health-risk. Correct private information is therefore a necessary condition for making choices optimal from the insured's perspective. Private information, however, can also lead to adverse selection in insurance markets, resulting in inefficient sorting or pooling equilibria, and even outright market breakdown (Rothschild and Stiglitz, 1976). This potential impact of private information on market functioning motivates the varying regulatory frameworks across healthcare systems protecting consumers and insurance providers (see e.g., Wuppermann (2017).

While variation in health beliefs unexplained by health information could indicate the presence of private information it could also be due to biased beliefs, thereby impeding decision making. Empirical research has found evidence for both hypotheses. Simple measures of self-assessed health have been shown to predict future health shocks up-and-above detailed health information (Nielsen, 2016; Wuppermann, 2017). Likewise, subjective mortality risk has been found to predict actual mortality conditional on objective risk factors (e.g. d'Uva et al. (2020)) suggesting that individuals have private information about their health. At the same time well-known cognitive biases such as overconfidence influence human decision-making across various domains (Zell and Krizan, 2014). The formulation of health beliefs and decision making on these is not exempt from this, with inaccurate health beliefs being found in large parts of the population and persistently associated with health prevention efforts (Arni et al., 2021; Spitzer and Shaikh, 2022).

The accuracy of private health beliefs therefore matters for the quality of economic decision making; biased health beliefs may lead to under (or over-) investments in ones' health, while private information on health care risk can lead to advantageous selection into health insurance. However, the relation between health beliefs and economic choices is complex as individuals might not know how to effectively use private health information. Also, (biases in) health beliefs might be correlated with preferences affecting decision-making like risk-aversion. Although the inability to use private information would lead to suboptimal decisions, when this private information is incorrect or leads to undesired selection effects, the welfare effects are not *a-priori* clear.

In this study, we investigate whether individuals' assessment of their own health has predictive power for their insurable health-risk and whether this information is used to make optimal

decisions when choosing health insurance coverage. For this, we exploit two advantages of the Dutch health insurance setting. Firstly, the availability of detailed administrative data on prior health care use and survey information on lifestyle, mental health, disabilities and self-reported health for over a million individuals. This allows us to control for a much richer set of objective health information, both when estimating or objective measure of health and when predicting the risk of health care costs. Secondly, the relative simplicity of the economic choice task of the Dutch health insurance deductible, allowing us to define the welfare optimizing choice relatively unambiguously.

We classify individuals into subjective health classes depending on whether they report a higher or lower general health than expected based on detailed health information mutually observed by insurers and insured and private health information. We then investigate whether the deviations from expected health have predictive power for the probability of having high health care costs and for the choice of the right deductible. We further investigate the relation between health beliefs and the ability to make optimal choices by comparing the deductible choice between individuals who have the same cost risk but different self-assessed health.

We contribute to three strands of literature. First, the literature on health belief formation and their relation with different dimensions of health information. While measures of perceived health are found to be generally correlated with objective health indicators, such as the existence of chronic conditions or risky health behaviors, there is considerable heterogeneity in these relationships across information dimensions and with respect to the strength of these relationships across subgroups (e.g., Jürges (2007) or Au and Johnston (2014)). A particular concern is the existence of biased beliefs, the systemic over- or under-estimation of one's health conditional on objective health information. Although studies have linked biases in, often quite particular, measures of health to health behaviors (Arni et al., 2021; Biró, 2016; Spitzer and Shaikh, 2022), there are few studies convincingly linking the relation between health perception and behavior to choice-quality and the resulting welfare effects. Studies doing so (e.g. Gan et al. (2015)) rely on relatively restrictive assumptions. Our main contribution here is that we link health perceptions to economic choices in a simple choice task in which the welfare implications of choices are much more straightforward,

Second, researchers have investigated the existence of private information in (health) insurance markets. Private information can result in adverse selection if individuals knowing more about their high risk than the insurer sort into insurance while individuals with low risk opt out. This could lead to a 'death spiral' where selection eventually leads to higher costs, leading to higher premiums, reinforcing selection (Rothschild and Stiglitz, 1976). Recent literature (e.g. Fang et al. (2008); Finkelstein and McGarry (2006); Finkelstein and Poterba (2006)) has shown that private information is just one dimension individuals sort on, and multidimensional sorting (e.g., on health and risk-aversion) sometimes prevents or worsens selection. Our set-up is quite close to the 'testing for adverse selection with unused observables' approach of Finkelstein and Poterba (2006), who consider whether there is information unknown to the insurer but known to the individual which is correlated with both outcome risk and insurance uptake. As most of these studies are interested in the difference between what the insurer and the

individual knows, and as insurers generally have limited information, they include only a few potential risk predictors. We include information (e.g. lifestyle) considered private by previous studies. Moreover, these studies tend to induce the existence of private health information from observed correlations between risk and insurance choice rather than from direct information on truly private information such as self-perceived health (Wuppermann, 2017). Or, they use very specific information such as the self-assessed risk to enter a nursing home (Finkelstein and McGarry, 2006).

Finally, there is an extensive literature on the quality of health insurance choices (see e.g., Bhargava et al. (2017)) with several papers focusing on the Dutch case. The first one is Handel et al. (2023) from which we borrow the idea of using differences in predicted risk and deductible choice to assess welfare effects. They use this approach to document socioeconomic differences in choice-quality: conditional on the same cost risk, individuals with lower socioeconomic status chose less optimally. Earlier work by Van Winssen et al. (2015) assessed the quality of the deductible choice by comparing the difference between the premium discount and ex-post out-of-pocket costs, unlike Handel et al. (2023) ignoring ex-ante risk-aversion. Two recent studies consider the relation between deductible choice, selection, and moral hazard. Remmerswaal et al. (2023) investigates the selection of low risk individuals into the high deductible. Exploiting the necessity to chose a deductible when turning 18, they show that low-risk individuals are more likely to select the high deductible, arguing that this this selection effect and not the reduction in moral hazard drives lower health expenditures among this group. Alessie et al. (2020) use a similar strategy as Finkelstein and McGarry (2006), using a small panel to conclude that private health information is predictive for costs and correlates with deductible uptake. They also find that conditional on health, individuals with the high deductible use less health care, which they interpret as evidence of moral hazard, contrary to Remmerswaal et al. (2023). In both studies, an implicit necessary link between individuals' risk of high health care costs and their choice is that they have private health information. We make this link explicit and disentangle individuals' private information from their ability to use it when choosing their deductible level.

II Institutional Setting and Optimal Insurance Choice

A The Voluntary Health Insurance Deductible

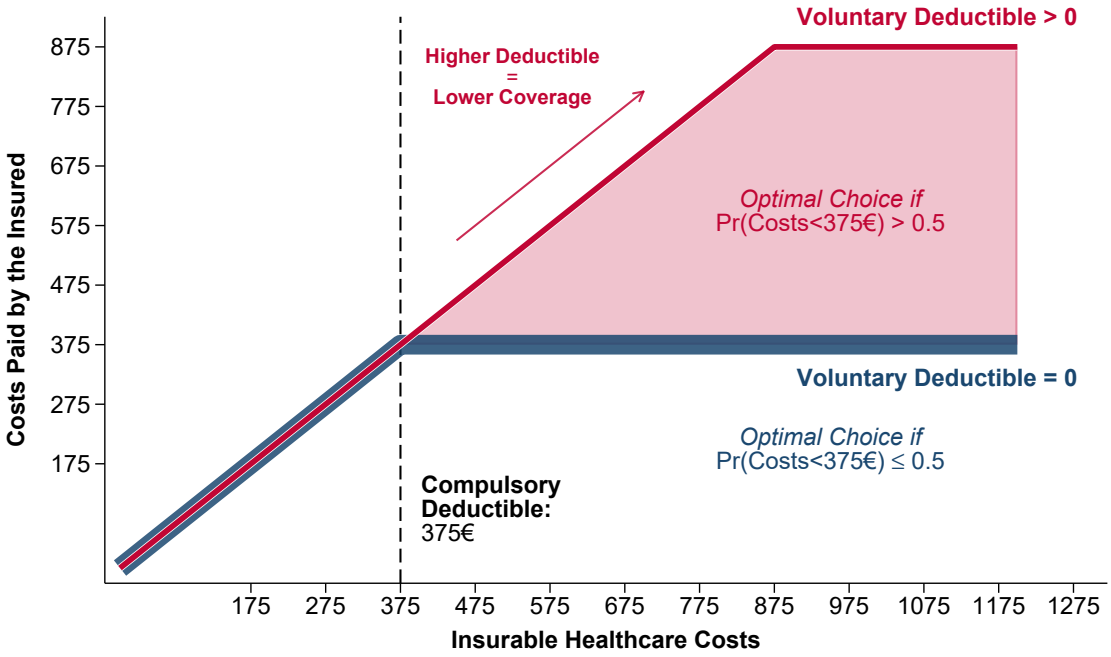
In the Netherlands basic health insurance is compulsory. There is competition between insurers but basic coverage levels are mandated by the government. Health insurers cannot risk-rate but a detailed compensating risk-equalization scheme exists. All insurance contracts include the compulsory deductible of currently €385 per year, but individuals can choose to increase this in steps of €100 up to €500 resulting in a maximum deductible of €885 (the compulsory €385 and the voluntary €500) for a premium discount. Insurance and deductible choice happens at the start of each year. The average annual premium for an insurance package with the low deductible is €1200. Around 9% of all insured opts for a voluntary deductible. Of these more than 70% percent, choose the maximum of €500 (Handel et al., 2023). The premium discount individuals receive for choosing the high deductible depends on the insurer with an average of €250 for those choosing the highest €500 deductible with little variation across insurers with most discounts between €200-€300 (Handel et al., 2023). Most costs are insurable with only certain parts of GP care, and all maternal care and nursing care being always insured.

B Optimal Choice in the Absence of risk-aversion

To illustrate we focus on the choice between the low deductible (€0) and the highest voluntary deductible (€500). In a standard expected utility framework, the choice-task at the start of the year is whether the expected utility with the low deductible (lower risk of out of pocket costs, but a higher premium) outweighs the expected utility with the high deductible (a lower premium, but a higher risk of out of pocket costs). If individuals are risk-neutral, the optimal choice depends on whether the expected payments (premium plus out of pocket costs) under the high deductible are lower than under the low deductible. This depends on the premium discount and the expected out-of-pocket costs under the high deductible. We follow Handel et al. (2023) and use the fact that if an individual does end up needing health care, the costs generally exceed €885. This means that individuals choosing the high deductible end up either having zero additional costs if under the limit of the low deductible or face the maximum possible €500 additional costs if exceeding that limit. Expected additional costs under the high deductible are then a function of the probability that an individual i will stay within the copayment limit: $\mathrm{E}(c_i) = \mathrm{P}(c_i > \tau) * 500$. As the premium discount received (€250) is half the level of the deductible (€500), an individual should choose the high deductible if $\mathrm{P}(c_i > \tau) < 0.5$, because then $\mathrm{P}(c_i > \tau) * 500 < 250$, and choose the low deductible if $\mathrm{P}(c_i > \tau) < 0.5$.

The attractive property of this framework is that the deductible choice is reduced to a straightforward choice-task with a clear-cut optimal strategy. As long as we have a credible estimate of $\mathrm{P}(c_i > \tau)$ (which potentially requires private health information), we can identify and compare optimal to actual choices. In case the actual choice deviates from the optimal, we individuals make a mistake (in the sense that their choice does not maximize their utility). This interpretation crucially depends on whether an individual's welfare is driven by the expected costs and not by other things, the most important of these other factors are risk preferences.

Figure 1: Deductible Choice Task and Optimal Decisions under Risk Neutrality



Source: Own illustration.

C Optimal Choice in the Presence of risk-aversion

We can generalize the decision problem allowing for risk-aversion by introducing a consumption utility function $U(.)$. In this case the optimal decision depends on whether expected utility under the high deductible, $\mathrm{E}(U_i^{high})$, exceeds that under the low deductible, $\mathrm{E}(U_i^{low})$. The utilities under the low and high deductible are:

$$\mathrm{E}(U_i^{low}) = U(y_i - p_{low}) \tag{1}$$

and

$$\mathrm{E}(U_i^{high}) = (1 - \mathrm{P}(c_i > \tau))U(y_i - p_{high}) + \mathrm{P}(c_i > \tau)U(y_i - p_{high} - o_{high}), \tag{2}$$

with y_i is income, p_{low} the fixed premium under the low deductible, $p_{low} - p_{high} = 250$ the premium discount under the high deductible, and $o_{high} = 500$ the additional out-of-pocket costs incurred under the high deductible when health care costs exceed the low deductible threshold.

Crucially, the optimal choice no longer depends only on $\mathrm{P}(c_i > \tau)$, but also on the functional form of the utility function. For instance, in a standard constant-absolute-risk-aversion (CARA) utility function, the probability at which the low deductible is optimal is lower than 0.5 and depends on the level of risk-aversion. Similarly, in case of constant-relative-risk-aversion (CRRA), the optimal probability of high costs to choose the low deductible depends on risk-aversion and income.

Handel et al. (2023) argue that for risk-averse individuals the optimal choice also only depends on whether the probability of exceeding the low-deductible threshold τ is larger than 0.5. This

is because the monetary amount involved are so small that the (second-order) effect of risk is of little importance. Using a constant absolute risk-aversion of 10^{-4} the high deductible is optimal when $\mathrm{P}(c_i > \tau) < 0.51$, while under a very high level of 10^{-3} choosing the high deductible is optimal if $\mathrm{P}(c_i > \tau) < 0.56$. In our main analysis, we follow this reasoning, interpreting differences between the actual choice and the optimal choice under risk-neutrality as sub-optimal from the insured's perspective.

III Data

A Administrative Data

We use administrative data provided by Statistics Netherlands. This data covers the period from 2009 to 2021, allowing us to observe individual-level longitudinal records of healthcare use and costs, income tax data, labour market participation, and demographic information.

Based on data from the national hospital admission register (LBZ), we identify whether an individual was admitted at least once to a hospital in a given year for each of the main chapters of the International Shortlist for Hospital Morbidity (ISHMT). Using data on annual prescription medication use we apply the encoding developed by Huber et al. (2013) and refined by Danesh et al. (2024) to infer the presence of 21 chronic conditions. Using electronic health claims data we measure healthcare costs by the cost-categories covering GP care, hospital care, expenses for prescription medications, mental health services, dental care, paramedical care (e.g., physiotherapy), maternal care, geriatric care and nursing care. This allows us to distinguish between uninsurable and insurable costs. Using data on household and individual-level income we measure 5-year age-gender specific income ranks based on per-capita household incomes of the entire population. Lastly, we observe chosen deductible levels for the Dutch population between 2011 and 2021.

B Survey Data

Our administrative data can be linked to survey datasets using pseudo-anonymized identifiers and we make use of two survey datasets. The first dataset, on the basis of which majority of our analyses are conducted, is the Dutch Health Monitor (*"Gezondheidsmonitor"*), from here on referred to as Health Monitor or GEMON, a large-scale cross-sectional health survey. The second dataset is the Longitudinal Internet studies for the Social Sciences (LISS), a longitudinal household panel survey covering around 5000 households between 2008 and 2023.

The Health Monitor is nationally representative health survey of the adult population in the Netherlands (aged 18+). It started in 2012 and is conducted every 4 years. Individuals registered in their respective municipality as living in a private household are invited to participate. Surveys are conducted between September 1^{st} and December 31^{st} by inviting individuals by letter to participate in the online survey[1]. We combine the 2012, 2016, and 2020 Health Monitors, yielding a raw sample of ≈1,19 million individuals identifiable in administrative records. While Health Monitor samples over-represent older individuals they are highly representative for the underlying Dutch population conditional on age. Figure A1.1 in the Appendix depicts the age-distribution and the share of individuals having a cardio-vascular disease, the share of individuals taking the high voluntary deductible, and the share of individuals with healthcare

[1] A paper-based questionnaire was included in some regions in the invitation letter while in others this was only sent to those not responding to the initial invitation. In-person or phone-based interviews make up only 0.5% (2012) and 0.1% (2016/2020) of collected responses. The 2012 Health Monitor was sent to approximately 700,000 individuals with 387,195 respondents (ca. 55%), while in 2016 1.15 million individuals were approached resulting in 457,153 respondents (ca. 40%) followed by 1.39 million individuals yielding 539,000 respondents (ca. 39%) in 2020.

costs above their copayment limit, comparing the entire adult Dutch population and the Health Monitor subsample.

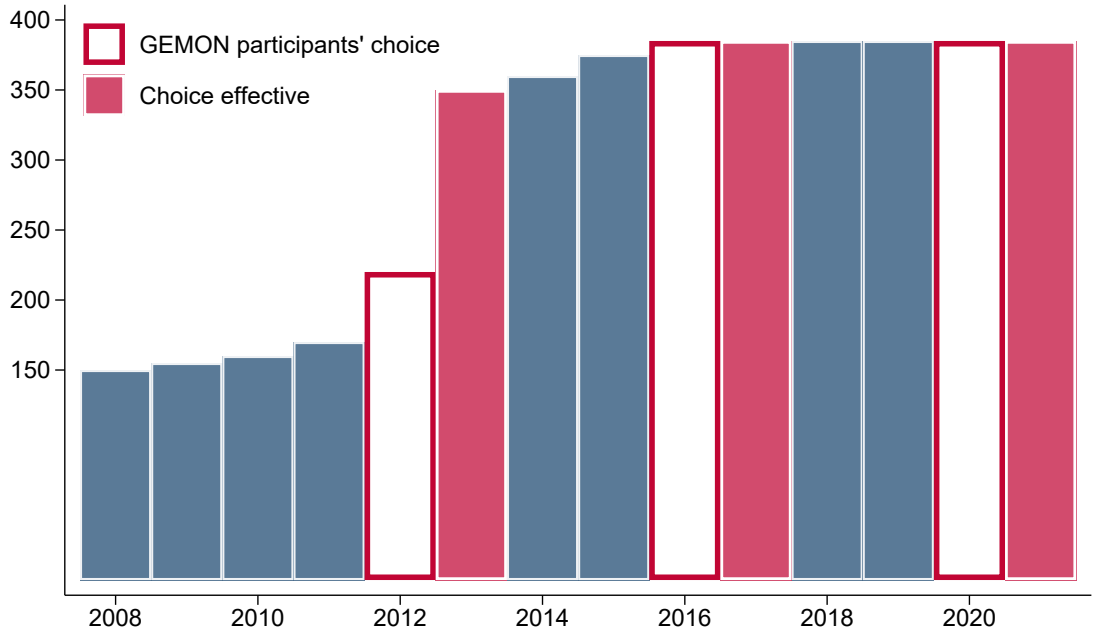
From the Health Monitor we include information on education status and self-reported information on health behaviors covering smoking (previous and current smoking status), weekly alcohol consumption, weekly exercise, and BMI. Further we include information on current health through seven dimensions of physical, cognitive, and sensory functional limitations (walking 400m without pause, lifting a 5kg weight and carrying it for 10 meters, being able to reach the ground, reading newspaper fine-print, hearing comprehension, the ability to hold a conversation, and facial recognition) which respondents indicate to be not limited due to their health, somewhat limited, limited a lot or entirely prevented to perform. For each dimension a dummy encodes respondents indicating to be limited a lot or entirely. Further we use the Kessler 10-item depression and anxiety questionnaire (Kessler et al., 1994), a clinically validated screening questionnaire for depression and anxiety related disorders. Lastly the Health Monitor contains a measure of perceived health in the form of self-assessed health. Individuals indicate on a five-point likert-scale ranging from *"Very Bad"* to *"Very Good"* their answer to the question *"Overall, how good would you say your health is?"*. While self-assessed health is a broad measure of individuals' health perceptions it has been shown to consistently be associated with future health outcomes (see e.g. Nielsen (2016)) and can be argued to be a valid measure of private health information precisely due to its non-verifiable and broad nature when combined with other health information (Wuppermann, 2017).

The LISS panel is a representative household panel survey covering similar variables as those included in the Dutch Health monitor alongside one-off survey modules on behavioural determinants of insurance choice such as risk-aversion or subjective survival expectations.

C Deductible Choice Timing and Analysis Samples

Figure 2 provides an overview of the compulsory health insurance deductible levels in the Netherlands between 2008 and 2021 and the timing of the observed deductible choices for each of the Health Monitor surveys. All individuals in the Netherlands need to chose their voluntary deductible anew at the end of a given year between mid-November and the end of December, a period in which insurers advertise their products and send reminders to their customers (Handel et al., 2023). The choice at the end of the year is effective for the subsequent year with the deductible being applicable to healthcare costs incurred in that year. If no choice is made the compulsory deductible level with a voluntary deductible of zero or is applied unless a deductible was chosen previously in which case this is the default. Given the timing of the Health Monitor surveys, which are conducted between September and December of a given year, we focus on the deductible choice made in the year of the Health Monitor participation (2012, 2016, and 2020) for the subsequent year (2013, 2017, and 2021) using observed healthcare costs incurred in that year to evaluate choice quality.

Figure 2: Compulsory Deductible Levels and Health Monitor Timing, 2008-2021



Source: Own calculations based on publicly available data.

We create two datasets. Firstly, a dataset of the entire Dutch population aged between 21-80 on January 1^{st} 2016 and surviving at least until January 1^{st} 2017, covering 11,547,759 individuals for which voluntary deductible levels in 2016 and administrative data for the two preceding years (2015 and 2014) are observed. Secondly, a pooled Health Monitor sample combining survey participants from 2012-2020 aged between 21-80 on January 1^{st} of the survey-year and surviving at least until January 1^{st} the subsequent year. Furthermore we condition on the availability of administrative records as for the general population dataset and the availability of all data from the Health Monitor survey as outlined above, yielding an analysis sample of 837,697 individuals.

IV Empirical Strategy

Our research question is whether private health information allows individuals to make optimal decisions when choosing their voluntary deductible, thereby improving choice-quality from the insured's perspective but resulting in adverse selection. Our empirical strategy contains three steps. Firstly we create a measure of health-risk based on information mutually observable by insurers and insured. The purpose of this is to obtain an individual-level health-risk measure at the time of the insurance choice as a benchmark against which added informational value of private information is compared (Section A). Following Handel et al. (2023) we use a cross-section of the entire adult Dutch population to calibrate age-gender specific health-risk models. In a second step we use the Health Monitor survey data to quantify the degree of private health information individuals have (Section B). Thirdly, we combine information on private and mutual health information and the predicted and actual realization of health outcomes and deductible choices to explore whether private health information allows individuals to make individually optimal choices (Section C).

A Population-based Health-Risk Model

In the first step of our analysis we create a measure of health-risk to capture the underlying risk relevant to the insurance choice we evaluate. As explained in Section B, individuals' optimal deductible choice depends (solely) on the risk of health care costs in year t exceeding the deductible threshold τ. From here onward we focus on the case of being an individual with costs below this threshold, specifically with costs below €375[2]. Whether considering the case of being high-cost ($c_i > \tau$) or low-cost ($c_i \leq \tau$) simply reverses the interpretation of observed empirical relationships and we focus on the low-cost perspective for ease of interpretation.

As we are interested in the added predictive value of private information an individual has beyond mutually observable health information, we first empirically estimate $\mathrm{P}(c_i \leq \tau)$ using predictors mutually observable by insurer and insured. Consider the vector of mutually observable health information $X_{M,i}$ in the years $t-1$ and $t-2$. In our context information on past healthcare costs by category, hospital care use by ISHMT group and the presence of chronic conditions . We use this data to estimate a linear probability model:

$$\mathrm{P}(c_i \leq \tau | X_{M,i}) = f(X_{M,i}). \tag{3}$$

As we have a large set of potential predictors whose importance is likely to vary across different groups and *ex-ante* not obvious interaction effects we use LASSO-shrinkage models estimated by 5-year age-gender groups to select the most predictive covariate compositions (Tibshirani, 1996), allowing for full interaction effects by age and gender. To make use of the broadest possible samples to identify the empirical relationship between $X_{M,i}$ and $\mathrm{P}(c_i \leq \tau)$ we use the general population dataset.

The resulting age-gender specific prediction models are then applied to the Health Monitor

[2]The compulsory deductible has varied over time but remained stable at €385 since 2016 (Figure 2) with €375 as the mean across this period .

survey participants. Using data on their mutually observable health information $X_{M,i}$ in the year of the survey participation $(t-1)$ and the year before $(t-2)$ we estimate the individual likelihood to have insurable healthcare costs below the copayment threshold $\hat{\rho}_M = \widehat{\mathrm{P}(c_i \leq \tau | X_{M,i})}$ in the subsequent years, 2013, 2017, and 2021 respectively.

B Measuring Private Information using Subjective Health Data

In the second step of our analysis we turn towards the Health Monitor survey data to create a measure of private health information *beyond* mutually observable health information $X_{M,i}$. We do this using data on self-perceived health in combination with information on privately held health information not observable by an insurer in the form of information on health behaviours, functional limitations, and mental health measures elicited in the surveys $(X_{P,i})$.

The creation of our private health information measure is straightforward. Let $H_i \in \{1, 2, 3, 4, 5\}$ be the subjective health variable, with $H_i = 1$ indicating the worst health state (*"Very bad"*) and $H_i = 5$ the best (*"Very good"*). We run linear regressions of the subjective health variable on the health-risk measure $\hat{\rho}_M$ capturing all mutually observable health information $X_{M,i}$ alongside the vector of private health information $X_{P,i}$:

$$H_i = \alpha_s + \beta_s X_{P,i} + \gamma_s \hat{\rho}_{M,i} + \varepsilon_i. \tag{4}$$

We use these linear regressions to identify variation in self-perceived health remaining unexplained by our measure of health-risk using mutual health information $(\hat{\rho}_{M,i})$ and the different dimensions of private health information $(X_{P,i})$. All regressions are run by 5-year age-gender groups as for our general population based health-risk model to allow for maximum flexibility in possible interaction effects. In all specifications we further control for single-year-of-age dummies and survey-year dummies. We then translate the unexplained variation in self-perceived health, the prediction errors of these regressions $H_i - \widehat{H_i} | X_{P,i}, \hat{\rho}_{M,i}$, into three subjective health classes $S_i \in \{1, 2, 3\}$:

- $S_i = 1$, Better health than predicted: $H_i - \widehat{H_i} | X_{P,i}, \hat{\rho}_{M,i} > 0.5$ [Higher SAH]
- $S_i = 2$, Equal health as predicted: $-0.5 \leq H_i - \widehat{H_i} | X_{P,i}, \hat{\rho}_{M,i} \geq 0.5$ [Neutral SAH]
- $S_i = 3$, Worse health than predicted: $H_i - \widehat{H_i} | X_{P,i}, \hat{\rho}_{M,i} < -0.5$ [Lower SAH]

These cut-off values are chosen as our underlying subjective health measure is a categorical variable where a residual exceeding 0.5 would indicate a deviation of one response category (e.g. from 4, *"Good"* to 3 *"Okay"*) if rounded to the nearest number[3]. Ultimately, however, these cut-off values are somewhat arbitrary. If individuals respond with subjective health assessments above or below what can be explained by mutually observable and privately known health characteristics this can either indicate that they have biased health beliefs or that they have additional private health information. In absence of an objectively true comparator we cannot formally separate the two. We can, however, assess whether the deviation in self-reported health from our health predictors contains relevant information for the insurable health-risk.

[3]Our results remain unchanged when instead using an ordered response model.

We chose this approach as it allows us to directly assess the importance of private information across individuals while also differentiating information contained in privately observed health characteristics $X_{P,i}$ from additional private information individuals have on their own health contained in H_i. It is common practice in studies considering the predictive power of private information for some outcome to simply add the subjective variable (in our case self-assessed health) as an additional predictor next to available information on mutually observed information and other privately observed characteristics (see e.g. Alessie et al. (2020); Finkelstein and McGarry (2006); Wuppermann (2017)). This way, it is possible to assess whether the private information contained in the subjective variable has additional predictive power *on average*. It does not, however, enable the identification of underlying heterogeneity: for some individuals all information might already be contained in the mutually or privately observable predictors, while others might have additional private information. Secondly, constructing a measure of private health information that is orthogonal to the objective information facilitates a more straightforward interpretation once we move to the cost-predictions: assessing the *added* predictive value of private health information contained in H_i over all other health information, we can simply look at the coefficient of our indicator encoding membership of the subjective health classes (S_i).

C Private Information and Insurance Choice

C.1 Private Information, Health-Risk and Deductible Uptake

In the third step of our empirical strategy we bring together the information on individual-level health-risk based on mutually observed health information with our measures of private health information. Combining data on health-risk predictions using mutually observed information $\hat{\rho}_{M,i}$ privately observed health characteristics $X_{P,i}$ and the additional unexplained variation in self-assessed health in the form of subjective health class membership S_i we run age-gender specific linear prediction models to determine individuals *ex-ante* likelihood of having low-costs:

$$\mathrm{P}(c_i \leq \tau | \hat{\rho}_{M,i}, X_{P,i}, S_i) = \alpha_c + \beta_c^1 \mathrm{I}[S_i = 1] + \beta_c^2 \mathrm{I}[S_i = 3] + \beta_c^P X_{P,i} + \gamma_c \hat{\rho}_{M,i} + \varepsilon_i \qquad (5)$$

We further considered two alternative specifications aggregating all information in the health-risk prediction. Firstly, by including the residuals from Equation 4 directly, thereby making use of the most granular variation in self-assessed health conditional on mutual and privately observed health characteristics. Secondly, by instead running the updated prediction models by subjective health classes (S_i) instead of including dummies, thereby allowing the coefficients for privately observed health characteristics to vary between the subjective health classes. All our results are consistent across these specifications and we present only the results from Equation 5.

Estimating Equation 5 allows us to explore two things. The updated individual-level health-risk prediction $\hat{\rho}_{MPS,i} = \mathrm{P}(c_i \leq \tau | \hat{\rho}_{M,i}, X_{P,i}, S_i)$, and the direct effect of membership in a specific subjective health class S_i on the likelihood of having low costs. To also consider how private health information influences deductible uptake we estimate:

$$\mathrm{P}(d_i = 1|\hat{\rho}_{M,i}, X_{P,i}, S_i) = \alpha_d + \beta_d^1 \mathrm{I}[S_i = 1] + \beta_d^2 \mathrm{I}[S_i = 3] + \beta_d^P X_{P,i} + \gamma_d \hat{\rho}_{M,i} + \varepsilon_i \qquad (6)$$

Where $d_i = 1$ indicates that an individual has chosen a deductible ≥ 0. If subjective health has predictive power for both the likelihood of being low-cost (Equation 5) and for the uptake of the deductible (Equation 6) this suffices to establish the presence of private information and resulting selection effects as we observe health-risk and insurance coverage (Finkelstein and Poterba, 2006). If better ($S_i = 1$) or worse ($S_i = 3$) than expected self-assessed health indicate a higher ($S_i = 1$) or lower ($S_i = 3$) likelihood to exceed the copayment threshold while also resulting in a higher ($S_i = 1$) or lower ($S_i = 3$) likelihood to take up the deductible we would observe adverse selection. Or the context of our decision problem, healthier individuals opt for lower coverage while less healthy individuals opt for high coverage. An opposing correlation on risk and choice we would indicate advantageous selection.

C.2 Private Information and Choice-Quality

While our empirical approach allows us to observe the presence of private information and establish whether adverse selection occurs it is difficult to assess on this correlation alone whether the behavioral response to the private health information is optimal. In the fourth analysis step we consider the interaction between health-risk and deductible uptake determining choice quality.

Under risk-neutrality individuals should choose a high deductible if their probability to be low-cost is higher than 0.5 and the low deductible only if this probability is lower than 0.5. We can use the predicted probability from Equation (5) together with observed deductible choices to measure choice-quality:

$$Q_i = d_i - \mathrm{I}[\widehat{\rho}_{MPS,i} > 0.5]$$

If $Q_i = 0$ an individual's choice is optimal, if $Q_i = -1$ an individual is over-insured (low deductible, while $\widehat{\rho}_{MPS,i} > 0.5$), and if $Q_i = 1$ an individual is under-insured (high deductible, while $\widehat{\rho}_{MPS,i} > 0.5$). To consider the role of subjective health classes on the quality of choices we run the following regressions:

$$\mathrm{P}(Q_i = c|\hat{\rho}_{M,i}, X_{P,i}, S_i) = \alpha_q + \beta_q^1 \mathrm{I}[S_i = 1] + \beta_q^2 \mathrm{I}[S_i = 3] + \beta_q^P X_{P,i} + \gamma_d \hat{\rho}_{M,i} + \varepsilon_i \qquad (7)$$

In addition we also run regressions for the broader case of being incorrectly insured irrespective whether it is over- or under-insured, ($Q_i \neq 0$. This focus on the *ex-ante* optimal choice allows us to also consider the heterogeneity of choice-quality *within* subjective health classes and across health-risk levels. Even among those who report better health than predicted, there are some with a high health-risk and similarly among those who report worse health there are individuals with a low health risk.

Focusing on the *ex-ante* choice-quality explicitly allows the quality of our health-risk predic-

tions to influence the observed choice-quality. In other words, our measure of choice-quality is determined by the influence of private health information in our updated cost-prediction model. An alternative approach is to instead consider not the *ex-ante* choice-quality but the *ex-post* realization of health-risk in correspondence with deductible levels chosen. We therefore create an alternative choice-quality measure that does not allow us to consider the underlying heterogeneity but is based on the actually observed interaction between d_i and $c_i \leq \tau$.

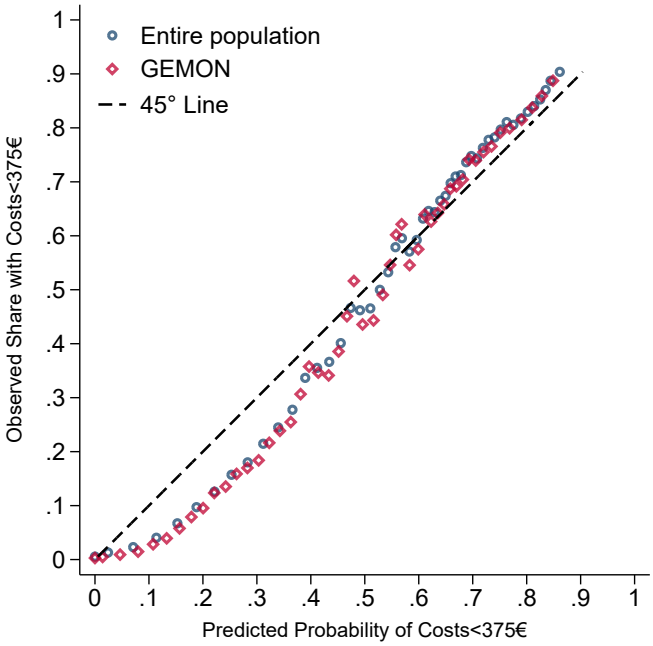
V Results

Our results are presented in line with the different steps of our empirical analyses. First we present evidence on the performance of our general health-risk prediction model using health information observed by insurer and insured (see Section A). Secondly we present results from our empirical models aimed at quantifying individual-level private information using survey data (see Section B) and whether the inclusion of this information adds relevant explanatory power with respect to health-risk and deductible (see Section C.1). We then present results on the quality of choices taking into account this private information (see Section C.2).

A Health-risk Prediction Model Performance

Figure 3 presents evidence on the performance of our cost-prediction models (see Section A) calibrated using the 2016 general population sample for the general population (blue dots) and our pooled Health Monitor sample (red squares). The y-axis depicts the observed share of individuals with costs below the copayment threshold (€375) while the x-axis depicts the predicted probability to have costs below the threshold based on observable mutual health information. The dashed diagonal line is a reference case in which our prediction model would yield predictions in which average predicted probabilities equal the actually observed share of individuals with low costs.

Figure 3: health-risk Prediction Model Performance



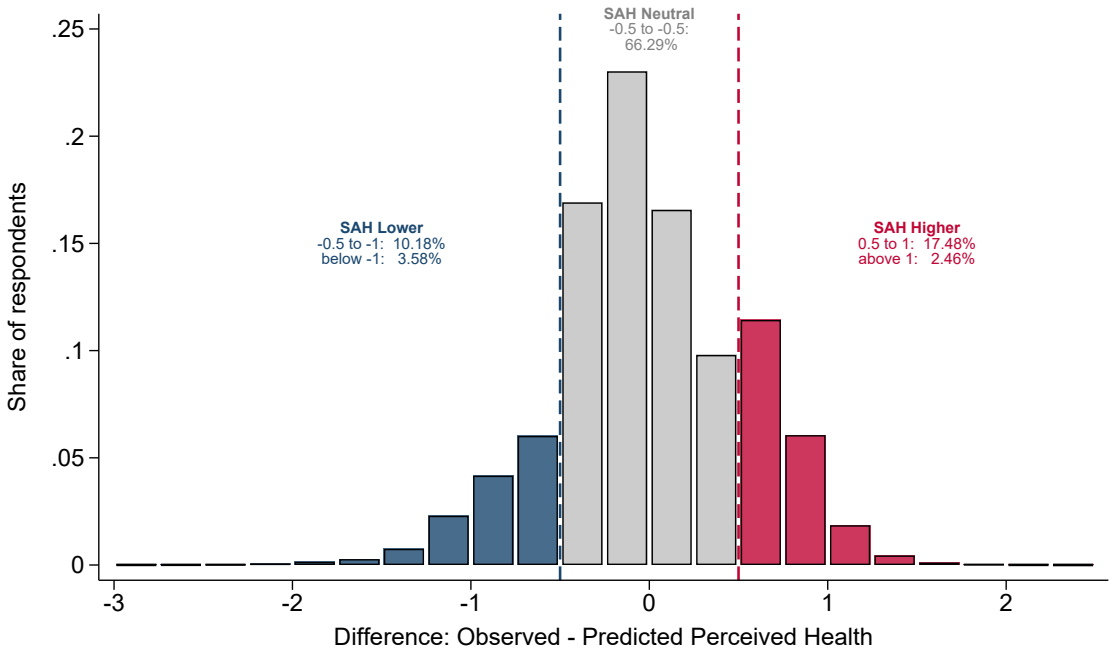
Source: Own calculations based on data from Statistics Netherlands. *Note:* The entire population sample contains all individuals aged between 20-80 on January 1^{st} 2016 (N=11,547,759) while the GEMON sample contains all unique participants of the Health Monitors 2012/2016/2020 aged between 20-80 on January 1^{st} of the respective survey year (N=837,697). Predicted probabilities come from the age-gender specific LASSO models using information on healthcare use in the two preceding years with potential predictors covering single-year-of-age-dummies, hospital care use, chronic conditions, and healthcare costs by category.

Overall our prediction model performs well. However, among those with higher health-risk (having a lower likelihood for low costs) our model tends to over-estimate while among those with lower health-risks it under-estimates. It is precisely this performance gap in which measure of private information would add additional explanatory power. At the same time, when considering the performance conditional on deductible choices (see Figure A1.2) we can observe that for those taking the high deductible the observed share of individuals with low costs is consistently above the predicted shares, providing further evidence on the importance of private information in explaining choice-quality.

B Private Health Information, Health-Risk and Deductible Choices

To explore the extend of private health information individuals have and whether this information offers additional explanatory power over mutually observable health information $X_{M,i}$ we estimate a measure of privately held health information as outlined in Section B. Figure 4 depicts the residuals from the regression of self-assessed health on the predicted probability of being low-cost and the private health information observed by individuals themselves (Equation (4)). The dashed vertical lines indicate the thresholds for the three different subjective health classes. Individuals on the left of the left (blue) line are classified as reporting subjective health lower than predicted, individuals on the right of the right (red) line as reporting a higher level than predicted. The broad majority of individuals' are classified as neutral with 68.29% of the sample falling into this category, 13.76% of individuals are classified as lower and 19.94% as higher.

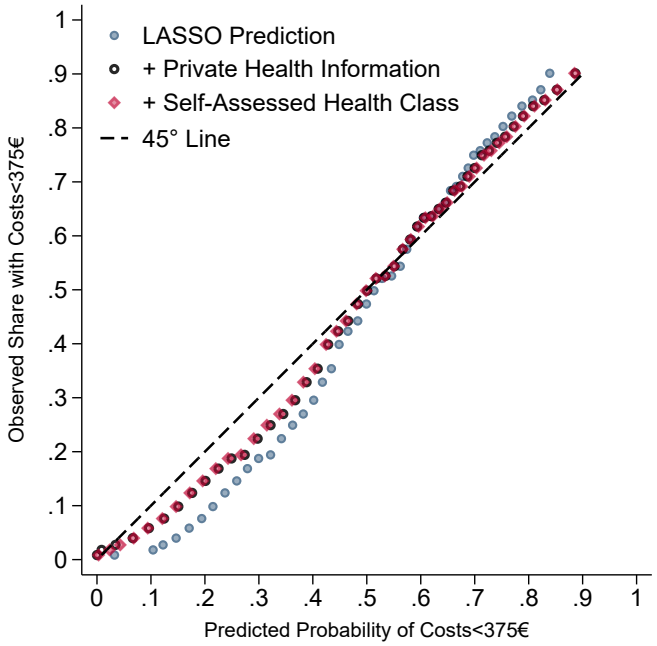
Figure 4: Self-Assessed Health Residual Distribution and Subjective Health Classes



Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the binned residuals from 5-year-age-gender group-specific regressions of self-assessed health on the predicted probability from the general population health-risk model alongside health information contained in the Health Monitor surveys (health behaviours and physical/mental health), single-year-of-age, and survey-year dummies using the pooled sample of Health Monitor participants (N=837,697).

Figure 5 depicts the performance of our health-risk prediction model when including private health information, contrasting the aggregate accuracy of our cost-predictions for the baseline health-risk model using only mutually observable health information $X_{M,i}$ (blue dots), against a model containing the predicted likelihood from this baseline model alongside private health information $X_{P,i}$ (black hollow dots), and a full model containing moth types of health information alongside dummies for SAH-class (red squares). As can be seen private health information not mutually observed improves the performance of our cost-prediction model considerably while the inclusion of SAH-class membership provides an additional but smaller improvement.

Figure 5: health-risk Prediction Model Performance with Private Health Information



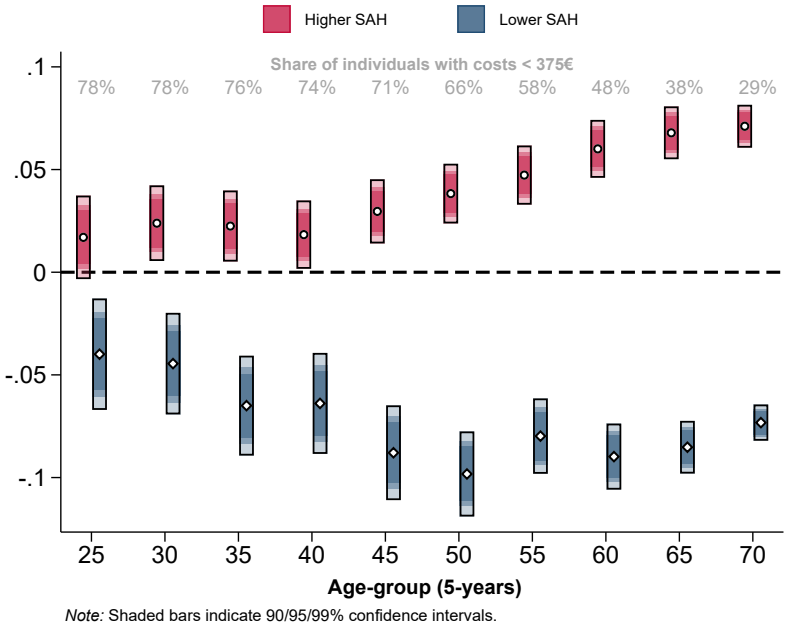
Source: Own calculations based on data from Statistics Netherlands. *Note:* Predicted linear probabilities come from 5-year-age-gender group-specific regressions of being low-cost on the predicted probability from the general population health-risk model alongside health information contained in the Health Monitor surveys (health behaviours and physical/mental health), single-year-of-age, and survey-year dummies using the pooled sample of Health Monitor participants (N=837,697).

To explore the role of private health information in our health-risk predictions and the deductible-uptake across different groups of individuals we can also plot the SAH-class regression-coefficients. Figure 6 plots the coefficients for males (panel a) and females (panel b) by age-group. The plots also depict the average share of individuals with costs below the copayment threshold as a benchmark for the interpretation of coefficient magnitudes. Across age-groups we observe that subjective health classes have distinct predictive power in line with what would be expected, individuals with lower subjective health than expected (blue) are less likely to have costs below the copayment threshold while individuals with higher subjective health (red) are more likely. Furthermore, as coefficients increase in older age-groups while the share of individuals with low costs decreases this relationship becomes increasingly stronger, suggesting private information is most relevant for older individuals.

When considering the role of private health information on deductible uptake the pattern as for costs replicates partially. Figure 7 depicts the coefficients from estimating Equation 6, focusing on the deductible choice, the effective insurance coverage individuals chose. While we observe a generally similar relationship with healthier-than-expected individuals (red) choosing higher coverage opposed to less-healthy ones (blue), the coefficients remain stable (males) or decrease (females) with age.

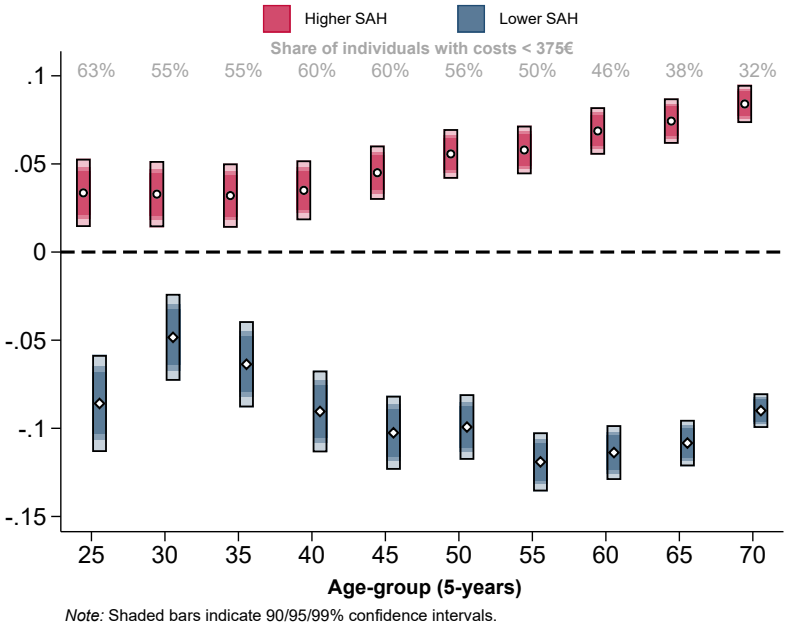
Figure 6: Subjective Health Classes and health-risk by Age

(a) Males



Note: Shaded bars indicate 90/95/99% confidence intervals.

(b) Females

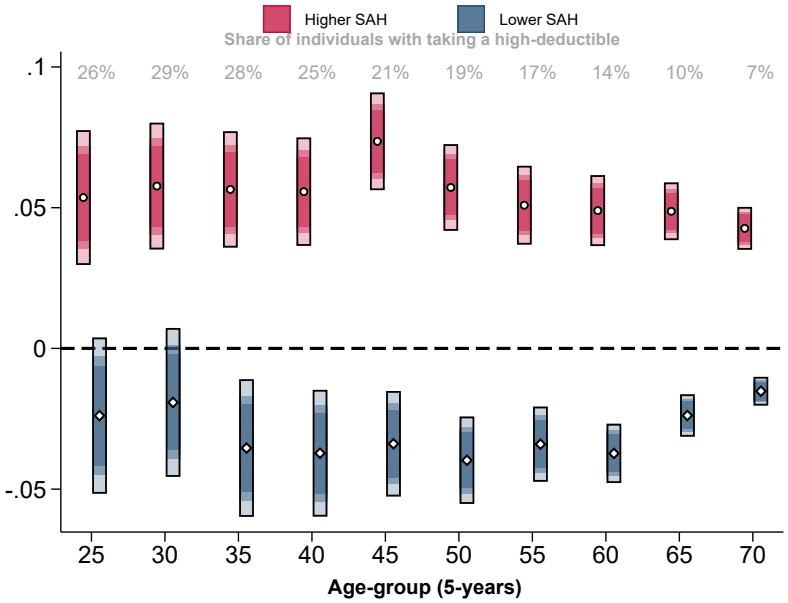


Note: Shaded bars indicate 90/95/99% confidence intervals.

Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the coefficients for subjective health class membership from 5-year-age-gender group-specific regressions of being low-cost in the year after the survey on the predicted probability from the general population health-risk model alongside health information contained in the Health Monitor surveys (health behaviours and physical/mental health), single-year-of-age, and survey-year dummies using the pooled sample of Health Monitor participants (N=837,697).

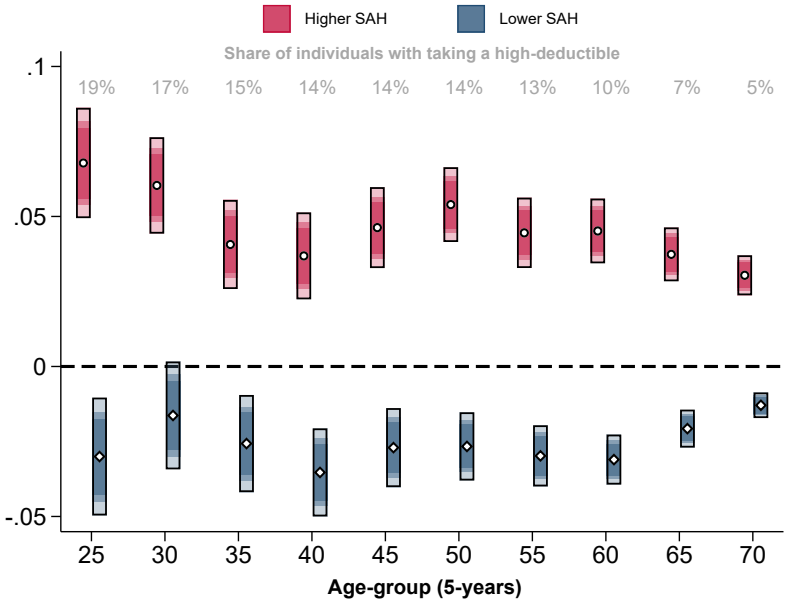
Figure 7: Subjective Health Classes and Deductible Uptake by Age

(a) Males



Note: Shaded bars indicate 90/95/99% confidence intervals.

(b) Females



Note: Shaded bars indicate 90/95/99% confidence intervals.

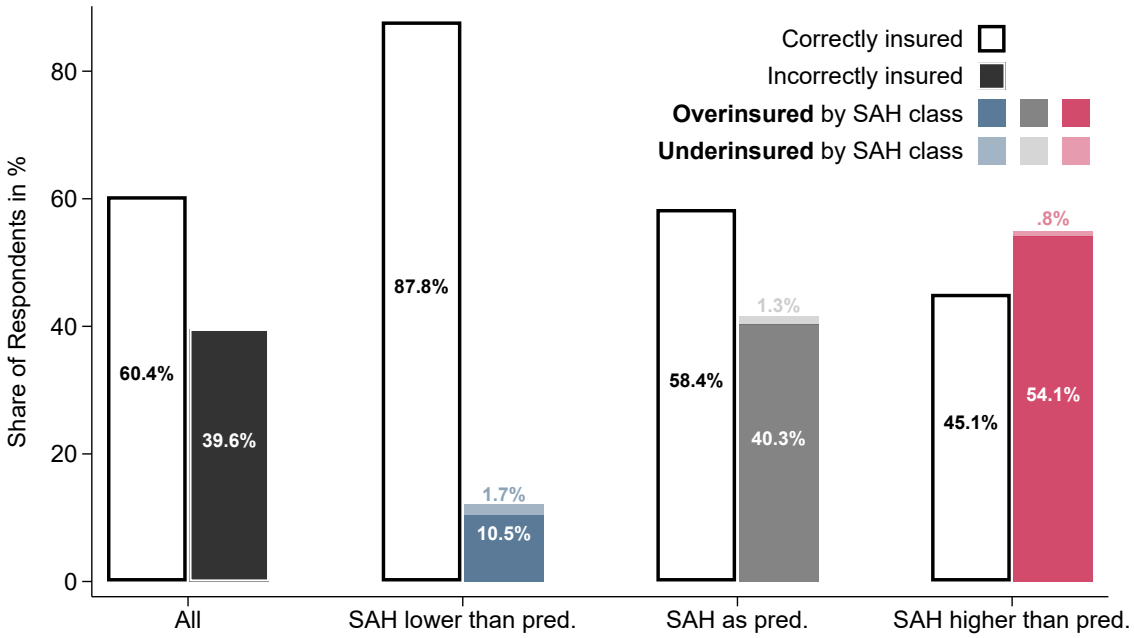
Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the coefficients for subjective health class membership from 5-year-age-gender group-specific regressions of taking a voluntary deductible in the year after the survey on the predicted probability from the general population health-risk model alongside health information contained in the Health Monitor surveys (health behaviours and physical/mental health), single-year-of-age, and survey-year dummies using the pooled sample of Health Monitor participants (N=837,697).

C Private Information and choice-quality

Our results show that individuals possess relevant private health information and use this information when choosing the deductible. As the observed coefficients from the regression on health-risk and deductible uptake suggest there is evidence for the effect on actual insurance choice to lag behind the predictive power in terms of health-risks. To what extent does this therefore impact choice-quality?

Figure 8 shows our main results regarding choice-quality by subjective health class. Overall, the majority of individuals chose a deductible level in-line with their predicted healt risk, with 60.4% being correctly insured. Of the 39.6% that are incorrectly insured the overwhelming majority, 38.4% percent of the population, are over-insured compared to a small population share of 1.2% being under-insured. Among individuals with lower than expected self-assessed health 87.8% chose the right deductible level in line with their health-risk , but this share is only 58.4% for those individuals with a neutral subjective health class. In stark contrast to this a minority among those with a higher subjective health class, only 45.1%, chose the right deductible. These results highlight the interaction between knowing about your health and optimally using this information and re-affirm the documented tendency among the Dutch population to avoid the high deductible (Handel et al., 2023).

Figure 8: choice-quality by Subjective Health Class



Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the share of respondents by *ex-ante* based choice-quality for the full sample (N=837,697) split into correctly and incorrectly insured, and each of the subjective health classes SAH higher (N=167,036), SAH neutral (N=555,394), and SAH lower (N=115,671) broken down into over-insured and under-insured.

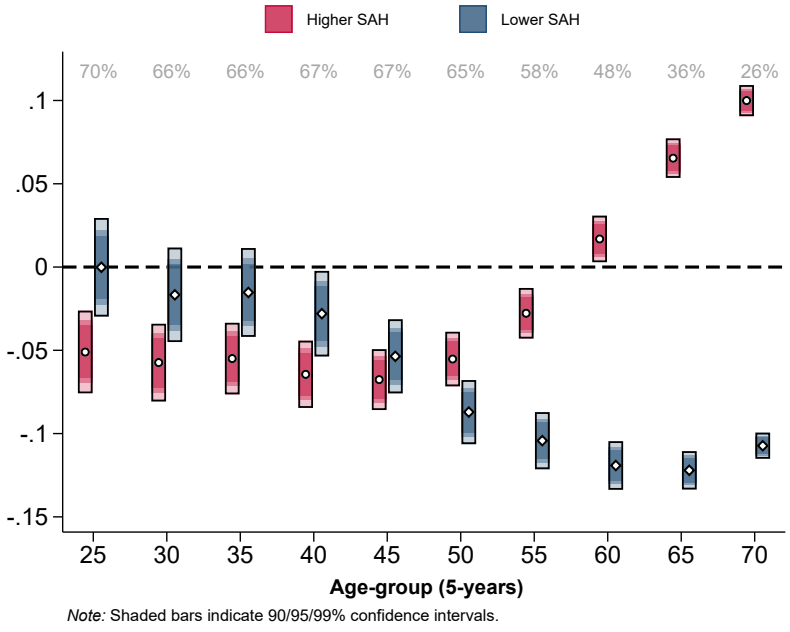
Turning to the impact of self-assessed health on choice-quality by age we present our results from estimating Equation 7 for the general case of being incorrectly insured (either over or

under-insured). Figure 9 depicts the subjective health class coefficients for our choice-quality measure using the *ex-ante* perspective. There is a clear age-specific pattern present for both male and female respondents, most pronounced among males. While being in the lower than expected subjective health class is associated with an increasingly lower likelihood for making an insurance error with increasing age the opposite holds for being in the higher than expected health class. Among these individuals for younger males it is initially associated with better choices before then becoming increasingly associated with making an insurance error. Given our results on the relationship between subjective health classes and the likelihood of incurring costs below the copayment threshold and the deductible uptake this is not surprising. The patterns of results on these outcomes already indicated deductible uptake lagging behind the predictive power individuals private health information contains. The results on our measure of choice-quality confirm that the implied pattern of average choice-quality also shown in Figure 8 does hold at the individual level and is driven by specific age-groups. When considering instead the *ex-post* measure of choice-quality we observe the same pattern of results albeit with generally smaller coefficients and wider confidence bands (Figure 10). The overarching pattern indicating the age-dependent insurance quality differential by subjective health class remains persistent also when taking this perspective.

Lastly we turn toward variation in choice-quality within the subjective health classes. To do so Figure 11 plots the share of individuals taking the high deductible on the y-axis against the predicted probability from our health-risk model using only mutual health information (Panel a) and using all available health information (Panel b). This figure provides two key insights. Firstly, the comparison of Panels a and b highlight again the importance of private health information. Once we take into account all available health information the gap in deductible uptake conditional on health-risk between subjective health classes becomes considerably smaller. Secondly, across subjective health classes choice-quality takes different forms. Individuals with higher than expected health while taking up the deductible with higher likelihood they do so insufficiently, in particular at the lower end of the health-risk distribution. This results in the lower overall choice-quality we observe in this group. However, individuals with a lower than expected subjective health class have high choice-quality precisely because they take up the deductible when in the lower end of the health-risk distribution and that is *despite the fact that their subjective health is lower than expected*. These results might indicate the importance of other over-riding choice determinants such as liquidity constraints. Alternatively these individuals might simply put a lower weight on their own subjective health as compared to other observable health characteristics.

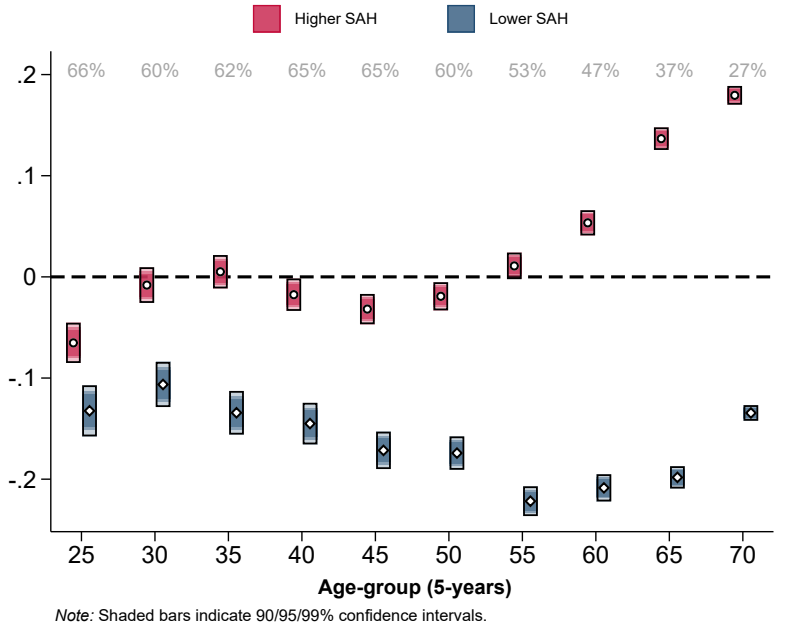
Figure 9: Subjective Health Class and *Ex-ante* Insurance-Errors by Age

(a) Baseline Predictions



Note: Shaded bars indicate 90/95/99% confidence intervals.

(b) Including Private Information

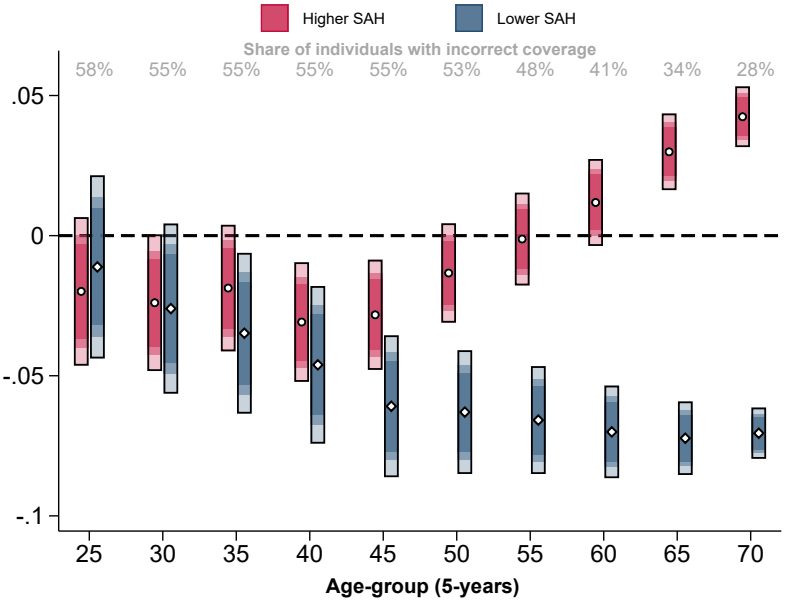


Note: Shaded bars indicate 90/95/99% confidence intervals.

Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the coefficients for subjective health class membership from a from 5-year-age-gender group-specific regressions of the *ex-ante* choice-quality measure indicating an incorrect deductible choice in the year after the survey participation on the predicted probability from the general population health-risk model alongside health information contained in the Health Monitor surveys (health behaviours and physical/mental health), single-year-of-age, and survey-year dummies using the pooled sample of Health Monitor participants (N=837,697).

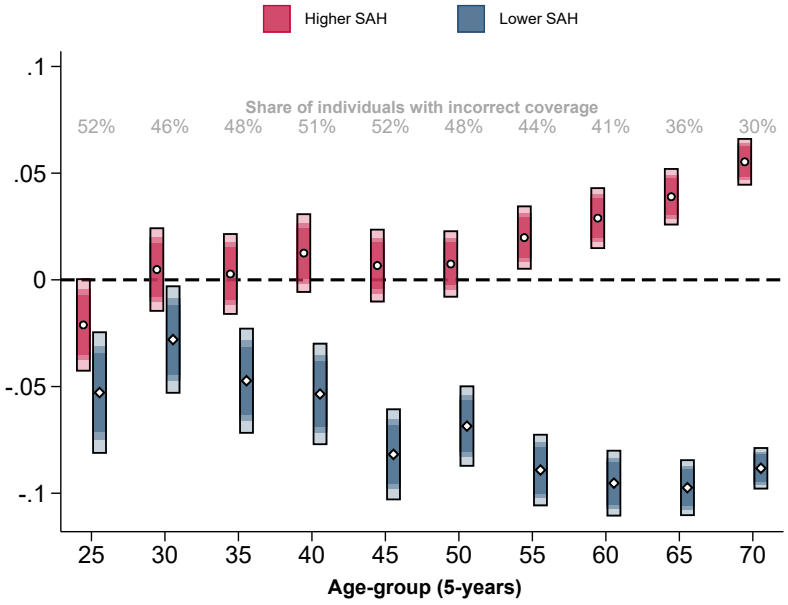
Figure 10: Subjective Health Class and *Ex-post* Insurance-Errors by Age

(a) Baseline Predictions



Note: Shaded bars indicate 90/95/99% confidence intervals.

(b) Including Private Information

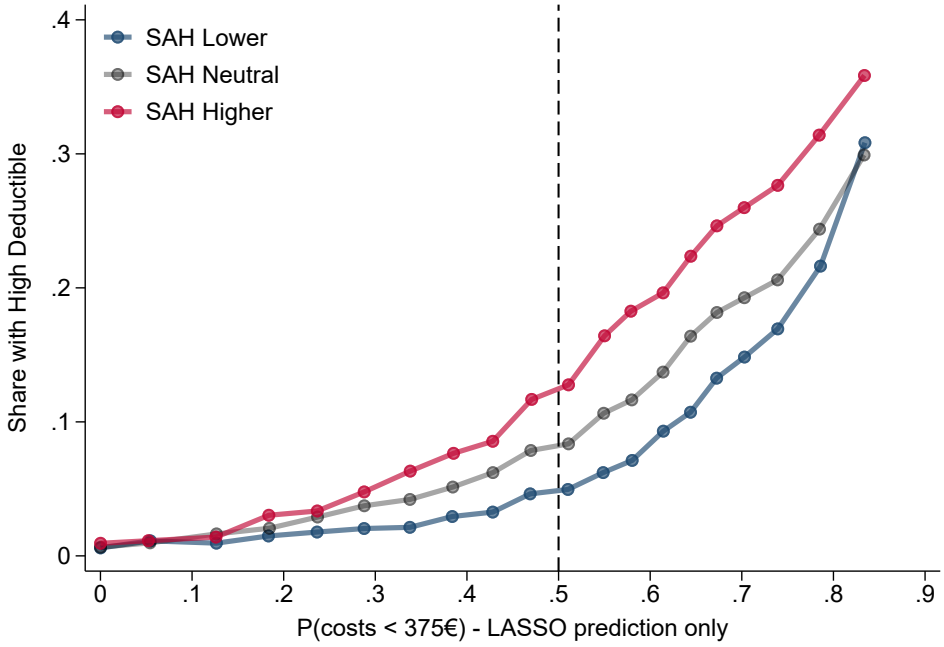


Note: Shaded bars indicate 90/95/99% confidence intervals.

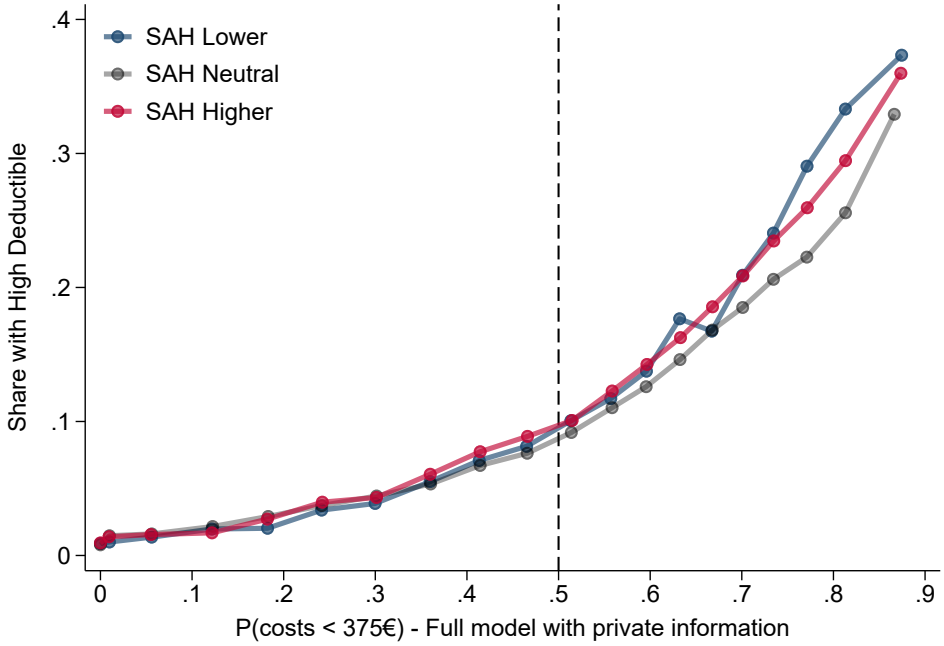
Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the coefficients for subjective health class membership from a from 5-year-age-gender group-specific regressions of the *ex-post* choice-quality measure indicating an incorrect deductible choice in the year after the survey participation on the predicted probability from the general population health-risk model alongside health information contained in the Health Monitor surveys (health behaviours and physical/mental health), single-year-of-age, and survey-year dummies using the pooled sample of Health Monitor participants (N=837,697).

Figure 11: Subjective Health Classes and Deductible Uptake by health-risk

(a) health-risk without Private Information



(b) health-risk with Private Information



Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the shares of individuals taking a deductible by subjective health class conditional on the estimated health-risk from our basic prediction model and our full prediction model using all available health information for the pooled sample of Health Monitor participants with N=837,697.

D Mechanisms and Additional Analyses

D.1 Socio-Economic Status and Education

As shown by Handel et al. (2023) the choice-quality of deductible choices in the Netherlands is depicts a socio-economic gradient. They find that individuals with higher incomes and better education levels overinsure on average but do so to a vastly lower extend. One potential explanation could be that education levels are correlated with the ability to make use of private information. To consider the role of both income and education we show the patterns of choice-quality ba health risk for different education and income levels. Figure A1.3 in the Appendix replicates Figure 11 but dividing the sample by subjective health classes and education and income levels. The figure replicates the results by Handel et al. (2023), showing a clear gradient in choice-quality especially by education and less so by income although. This provides some tentative evidence that education might play a role in the degree to which individuals make use of the private information they have while income levels seem to only be associated with an overarching shift of the deductible uptake curves. In particular with respect to the lower take-up of the deductible among lower income individuals this could also be explained by budget constraints or relative risk-aversion.

D.2 risk-aversion

Another potentially salient driver of deductible choices and the degree of over-insurance we observe is risk-aversion. As outlined given the small monetary amount involved in the decision task absolute risk-aversion would need to be very high to explain the observed uptake patterns (Handel et al., 2023). We generally do not observe risk-aversion in large data sets, but there is evidence from smaller surveys that health and health behaviors can be correlated with risk-aversion (Finkelstein and McGarry, 2006). The fact that the economic choice task at hand does depend on health-related risk but not on risk-aversion, means that we can be relatively confident that any differences in deductible choice we find across health classes are not caused by potentially correlated differences in risk-aversion.

To directly test whether risk-aversion does not play a role we can make use of supplementary data in the form of the LISS household panel survey which contains a module on risk-aversion. Participants were presented with a set of five choices in which they could choose between a fixed gamble with an expected value of €35 and varying guaranteed payoffs into a simple measure of risk-aversion by summing up the number of sure choices made (Alessie et al., 2020). We pool all LISS waves between 2011 and 2018 for which the same information on objective health information and private health information and subjective health perceptions is available. This results in a sample of around 4000 individuals and 15660 individual-year observations. The LISS survey covers all variables necessary to replicate the analyses done with the Health Monitor sample, albeit with a considerable lower number of observations requiring us to do so on a pooled sample of all respondents. We find that risk-aversion levels are entirely uncorrelated with our measure of private information, subjective health classes. While we observe generally that risk-aversion strongly determines deductible uptake in general it does not impact deductible uptake *conditional* on individual level health-risk and subjective health class membership. Given that

our measure of risk-aversion is based on a set of gambles individuals took with very small pay-offs we are fairly confident that this is not due to us looking at a measure of absolute risk-aversion that might not adequately reflect individual-level relative risk-aversion.

VI Discussion & Conclusion

In this paper we provide evidence on the role of private health information in determining insurance choice-quality. Combining large-scale survey data with administrative records we analyzed the quality of health insurance deductible choices in the Netherlands to disentangle the welfare implications of private health information due to its impact on choice quality.

Our empirical results provide three insights into the role of private health information in determining choice-quality. Firstly, we show that private health information is considerable and that even when operationalized using a general measure of subjective health it retains distinct predictive information beyond mutually observable health information. Importantly, while our analysis focused on subjective health as a measure of private information our results provide a lower-bound estimate given that we also control for a range of privately observed health characteristics. Secondly, individuals do make use of this information when choosing their coverage but do so to a far lesser degree than would be optimal. Thirdly, due to this overall choices are far from optimal. We find that individuals with private information indicating a low risk of future healthcare needs tend to be over-insured.

These findings imply that establishing that individuals have private health information and use this information in decision making is not sufficient to establish the welfare effects, and thus optimal policies. Quite some research has documented that private health information can be predictive for economic outcomes and that individuals tend to use this information, in choosing health insurance (see e.g. Finkelstein and McGarry (2006) or Wuppermann (2017)) but also for instance in making decisions about preventive care use (Fadlon and Nielsen, 2019). However, having information and using it in a decision is not equivalent to using it optimally as individuals can over- or under-respond to their information.

Our results also have implications for the design of the deductible. First, Remmerswaal et al. (2023) have raised the concerns that low-risk individuals select into choosing the high deductible, resulting in the high deductible function as a way to opt-out of the mandatory solidarity within the Dutch health insurance system. However, the risk of positive self-selection into high deductibles seems limited in practice and thus the welfare loss may be limited too. We find that most individuals, even those who have private information on being low-risk, opt for the low deductible. In other contexts, for example the reaction to ambiguous health information, the inability of individuals to correctly use private information can result in costly consumption of low-value care (Hoagland, 2021). We show that similar frictions can result in individually sub-optimal decision patterns beneficial for the system as a whole.

Secondly, Handel et al. (2023) have documented socioeconomic inequalities in deductible choice quality. They find that, mainly, individuals with a lower socioeconomic status with a low risk tend to stick to the low deductible and are thus over-insured. It is however unclear what drives this inequality: the ability to use private information or risk-aversion. We do find evidence suggesting that individuals with higher education make generally more use of the deductible and also do so once we incorporate private information in their risk predictions and condition on the same risk levels. However, over-insurance remains high conditional on having private

information irrespective of education or income levels.

A Planned Extensions

Our study has limitations guiding our ongoing work on future extensions. Firstly, the correct prediction of health-risks is a central component of our empirical strategy. While our current prediction models perform well there is room for improvement. We are considering to fully replicate the ensemble machine learning model used by Handel et al. (2023) while also introducing a Bayesian updating based adjustment of any general population risk model for our survey samples. If our results persist even if a health-risk prediction model yields on average perfect predictions this would provide strong evidence for relevant private information.

While we find little evidence for risk-aversion as the main driver of our results on over-insurance we plan to extend these by also considering relative risk-aversion. Further, other explanations exists that we can explore. Individual-level liquidity constraints, inertia or reverting to default choices and the role of survival expectations might at least in part explain our results. The former two we can directly explore based on available information on individual-level assets and the history of deductible choices, allowing us to distinguish never-switchers or those who might be driven by social network effects to change their deductible (see e.g. Handel et al. (2023)). Using data from LISS on the other hand we can directly incorporate survival expectations in our analyses, albeit only for a small sub-sample of our population. Data on survival expectations would also allow us to consider the role of general decision biases like overconfidence as we can observe objective benchmarks for these subjective assessments, unlike for self-assessed health. Lastly the health monitor surveys contain information on locus of controls, a behavioral determinant that has been linked to the demand for health insurance (Bonsang and Costa-Font, 2022) and has been found to moderate individual-level behavioural responses to health events (Schurer, 2017).

B Conclusion

In summary our study shows that individuals possess diverse private health information. Even when accounting for behavioural risk factors or other private health characteristics, a simple measure of subjective health still contains relevant information on future health-risks. Private health information is particularly informative for older individuals but it is also precisely among these individuals that we observe an insufficient usage of this information resulting in over-insurance. This suggests that while private information might remain relevant even if insurers can use detailed individual level information in the underwriting process (Wuppermann, 2017) in theory the threat of adverse selection remains. However, knowing ones health and acting upon this knowledge are two separate things. Individuals tend to make suboptimal decisions by not using the available information they have to the fullest degree.

References

Alessie, R. J., Angelini, V., Mierau, J. O., and Viluma, L. (2020). Moral hazard and selection for voluntary deductibles. *Health Economics*, 29(10):1251–1269.

Arni, P., Dragone, D., Goette, L., and Ziebarth, N. R. (2021). Biased health perceptions and risky health behaviors—theory and evidence. *Journal of Health Economics*, 76:102425.

Au, N. and Johnston, D. W. (2014). Self-assessed health: what does it mean and what does it hide? *Social science & medicine*, 121:21–28.

Bhargava, S., Loewenstein, G., and Sydnor, J. (2017). Choose to lose: Health plan choices from a menu with dominated option. *The Quarterly Journal of Economics*, 132(3):1319–1372.

Biró, A. (2016). Differences between subjective and predicted survival probabilities and their relation to preventive care use. *The BE Journal of Economic Analysis & Policy*, 16(2):807–835.

Bonsang, E. and Costa-Font, J. (2022). Buying control¿locus of control'and the uptake of supplementary health insurance. *Journal of Economic Behavior & Organization*, 204:476–489.

Danesh, K., Kolstad, J. T., Spinnewijn, J., and Parker, W. D. (2024). The chronic disease index: Analyzing health inequalities over the lifecycle. Technical report, National Bureau of Economic Research.

d'Uva, T. B., O'Donnell, O., and Van Doorslaer, E. (2020). Who can predict their own demise? heterogeneity in the accuracy and value of longevity expectations. *The Journal of the Economics of Ageing*, 17:100135.

Fadlon, I. and Nielsen, T. H. (2019). Family health behaviors. *American Economic Review*, 109(9):3162–3191.

Fang, H., Keane, M. P., 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. (2006). *Testing for Adverse Selection with" Unused Observables"*. National Bureau of Economic Research.

Gan, L., Gong, G., Hurd, M., and McFadden, D. (2015). Subjective mortality risk and bequests. *Journal of Econometrics*, 188(2):514–525.

Handel, B., Kolstad, J., Minten, T., and Spinnewijn, J. (2023). The socio-economic distribution of choice quality: Evidence from health insurance in the netherlands. *American Economic Review: Insights*.

Hoagland, A. (2021). An ounce of prevention or a pound of cure? the value of health risk information.

Huber, C. A., Szucs, T. D., Rapold, R., and Reich, O. (2013). Identifying patients with chronic conditions using pharmacy data in switzerland: an updated mapping approach to the classification of medications. *BMC public health*, 13:1–10.

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

Kessler, R., Mroczek, D., et al. (1994). Final versions of our non-specific psychological distress scale. *Ann Arbor, MI: Survey Research Center of the Institute for Social Research, University of Michigan*.

Nielsen, T. H. (2016). The relationship between self-rated health and hospital records. *Health Economics*, 25(4):497–512.

Remmerswaal, M., Boone, J., and Douven, R. (2023). Minimum generosity levels in a competitive health insurance market. *Journal of Health Economics*, 90:102782.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, 90(4):629–649.

Schurer, S. (2017). Bouncing back from health shocks: Locus of control and labor supply. *Journal of Economic Behavior & Organization*, 133:1–20.

Spitzer, S. and Shaikh, M. (2022). Health misperception and healthcare utilisation among older europeans. *The Journal of the Economics of Ageing*, 22:100383.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

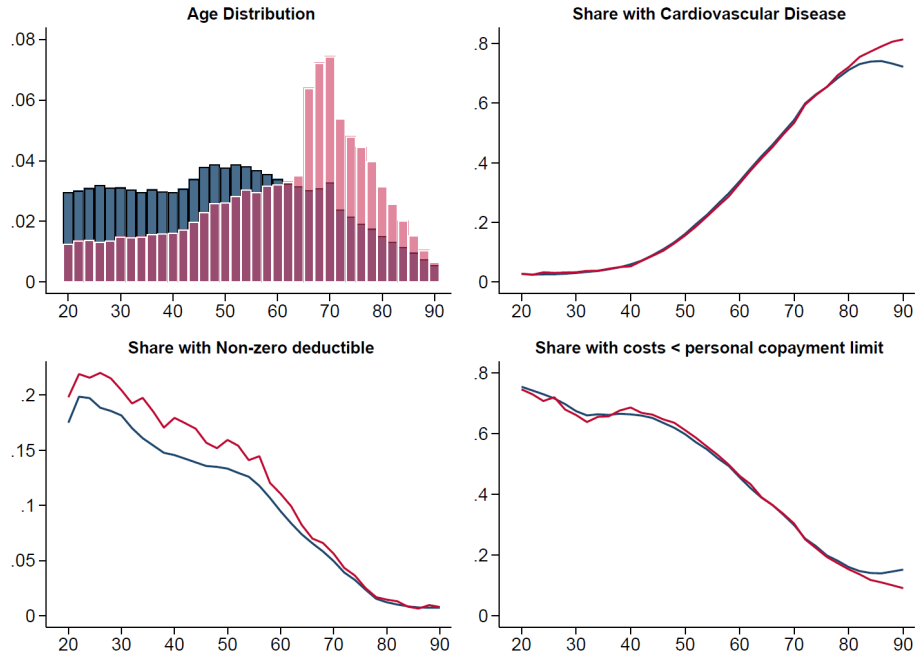
Van Winssen, K., Van Kleef, R., and Van de Ven, W. (2015). How profitable is a voluntary deductible in health insurance for the consumer? *Health Policy*, 119(5):688–695.

Wuppermann, A. C. (2017). Private information in life insurance, annuity, and health insurance markets. *The Scandinavian Journal of Economics*, 119(4):855–881.

Zell, E. and Krizan, Z. (2014). Do people have insight into their abilities? a metasynthesis. *Perspectives on Psychological Science*, 9(2):111–125.

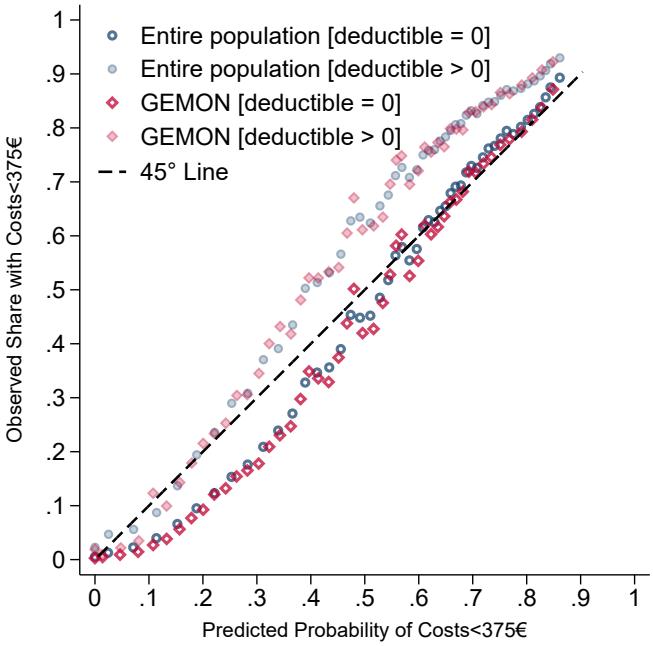
A1 Additional Tables and figures

Figure A1.1: General Population and Health Monitor Characteristics



Source: Own calculations based on data from Statistics Netherlands. Depicted is data from 2016 with blue bars/lines indicating data from the underlying Dutch population including the Health Monitor 2016 participants and red bars/lines correspond to the Health Monitor participants alone. The comparison of underlying populations to Health Monitor participants is similar for the years 2012 and 2020.

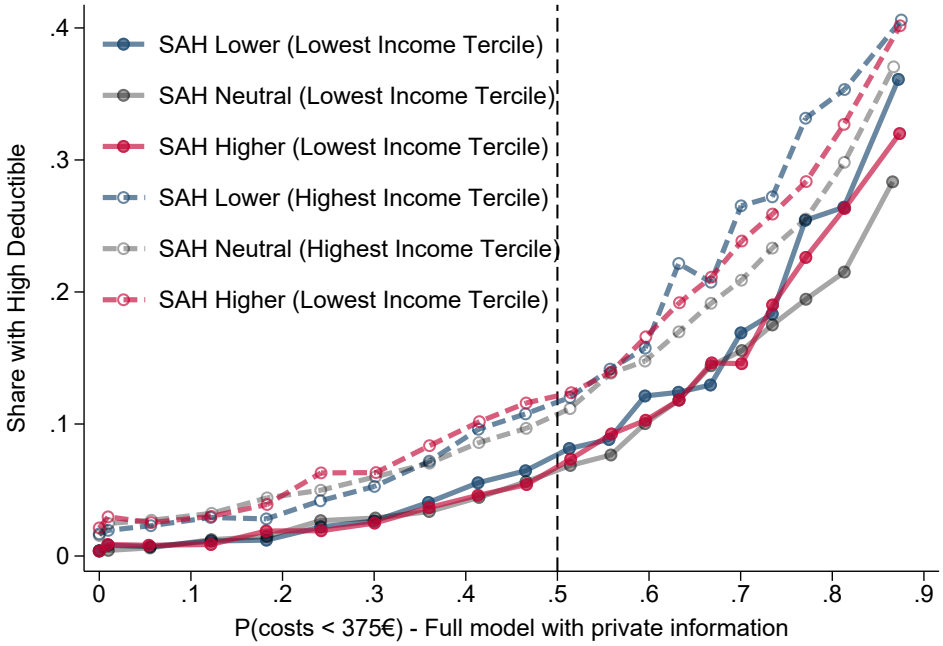
Figure A1.2: Health Risk Prediction Model Performance by Deductible Choice



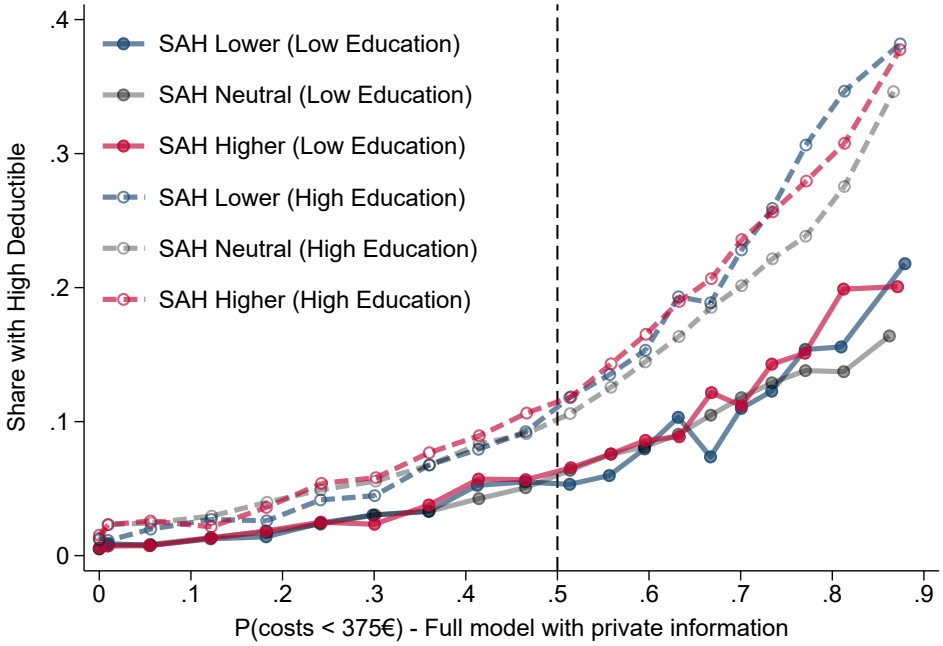
Source: Own calculations based on data from Statistics Netherlands. *Note:* The entire population sample contains all individuals aged between 20 and 80 years on January 1st 2016 (N=11,547,759) while the GEMON sample contains all unique participants of the Health Monitors 2012/2016/2020 aged between 20 and 80 on January 1st of the respective survey year (N=837,697).

Figure A1.3: Subjective Health Classes and Deductible Uptake by Health Risk and Education/Socio-Economic Status

(a) Income Terciles



(b) Education



Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the shares of individuals taking a deductible by subjective health class and broken down into income groups (by tercile) and education groups () conditional on the estimated health risk from our basic prediction model and our full prediction model using all available health information for the pooled sample of Health Monitor participants with N=837,697.