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

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

Tuesday 7th January, 2025

[Do not cite or circulate without consent]

Abstract

Individuals' perceptions of their own health might deviate from what can be observed based on objective health indicators, either because of biased health beliefs or because of private information. In both cases, whether individuals also act based on their private beliefs matters. It is often hard to distinguish between what individuals think about their health and how they act upon it. We combine large-scale survey data with granular administrative data and analyze a straightforward insurance choice. We study the deductible choice for Dutch adults, who have to take out a compulsory health insurance each year. They can choose between a low and a high deductible. The optimal choice between the low and high deductible depends on the expectation of exceeding the deductible. We link three waves of the Dutch Health Monitor, a health survey with more than 1 million adult participants between 2012 and 2020, with individual-level administrative data on past and future health outcomes. We distinguish between three sources of health information: (1) data on past health care use (2) information on health behaviors (3) additional private information. We estimate two sets of models. First, we estimate the probability of exceeding the deductible using only information on past health care use, and subsequently adding health behaviors and self-assessed health. Second, we do the same for take-up of the high deductible. The former allows us to identify whether individuals have relevant private health information, the latter whether they use this information to make better choices. Our results show that individuals' beliefs about their health are predictive for health care costs, even when controlling for extensive administrative information on prior health care use and self-reported lifestyle behavior. This private information is also used by individuals when they choose their health insurance deductible. However, our results also show that private information alone is not sufficient: many people, especially older healthier individuals, who have (private) information do not make deductible choices in line with this information.

Keywords: Health insurance, private information, health perceptions, choice quality.

JEL Classification: H00, I13, G52, D81, D90

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

1 Introduction

Do people have relevant private information and use this information effectively when making economic decisions? These two related but separate questions together determine the welfare effects of private beliefs in economic contexts, including insurance choice. We study both these questions in a setting that allows us not only to establish whether there is a correlation between having private information and the choices made but also assesses the effects of private information on individual welfare by considering the quality of choices made. To do so we utilize large-scale general population survey data and detailed administrative records on health outcomes and insurance choices to study the role private health beliefs play in explaining the quality of deductible choices in the Dutch health insurance system.

Individuals may think differently about their own health than what would be predicted based on health information observable by insurers. Correct private information allows individuals in principle to chose their insurance coverage optimally in line with their individual health risk exposure. From an individual's perspective correct private information is therefore a necessary condition for making optimal choices. Private information, however, is also a necessary condition for 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 the functioning of health insurance markets motivates the varying regulatory frameworks across healthcare systems designed to protect both consumers and insurance providers (see e.g., Wuppermann (2017).

While variation in health beliefs unexplained by observable health information could indicate the presence of private information it could also be due to incorrect or biased health beliefs, thereby impeding individuals' optimal decision making. Empirical research has found evidence for both of these things to occur. Simple measures of self-assessed health have been shown to predict future health shocks up-and-above other detailed health information (Nielsen, 2016; Wuppermann, 2017). Likewise, subjective information on mortality risk has been found to predict actual mortality even conditional on objective risk factors (e.g. d'Uva et al. (2020)) suggesting that individuals have relevant 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 overconfidence in ones health being found in large parts of the population and persitently associated with lower levels of 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 can be complex; making economic choices can be complicated and individuals might not always know how to use private health information to make an optimal decision. Also, (biases in) health beliefs might be correlated with preferences like risk-aversion which also affect decision making. Although generally

an inability to use private information in economic decision making would lead to suboptimal decisions, when this private information is incorrect or leads to undesired selection effects, the welfare effects of not fully using this information are not a-priori clear either.

In this study, we investigate whether individuals' assessment of their own general health has predictive power for their risk of high health costs and whether this information is used to make optimal decisions about health care insurance. For this, we exploit two main advantages of the Dutch health insurance setting. The first one is 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. The second one is the relative simplicity of the economic choice task, which allows us to define the welfare optimizing choice relatively unambiguously for each individual and compare this to the actual choice.

We classify individuals into subjective health classes depending on whether they report a higher or lower general health than can be expected based on our detailed health information covering extensive health records mutually observed by insurers and the insured and private information on additional health characteristics. 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 high 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, there is a long-standing literature on the formation of health-perceptions and their relation with different dimensions of health information. While measures of perceived health are found to be generally highly correlated with objective health indicators, such as the existence of a certain condition or health behavioural risk factors, there is considerable heterogeneity in the relationship between specific dimensions of objective health information and perceived health and with respect to the strength of these relationship across subgroups (e.g.Jürges (2007) or). A particular concern is the existence of health perception biases, the systemic over- or under-estimating 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 that convincingly link the relation between health perception and behavior to optimal choices with a clear-cut translation into welfare effects. Studies that do so (e.g. Gan et al. (2015)) tend to rely on relatively restrictive assumptions. Our main contribution here is that we are able to link health perceptions to economic choices in a simple choice task, in which the implications for optimal choice (and welfare) is much more straightforward,

Second, researchers have investigated the existence of private information in the (health) insurance market to address issues of adverse selection. In the traditional one-dimensional view on the relation between private information and insurance choice, there is the fear that individuals who know 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, again leading to more selection and so forth. More recent literature (e.g. Fang et al. (2008); Finkelstein and McGarry (2006); Finkelstein and Poterba (2006)) has shown that private information is just one of the dimensions individuals might sort on, and multidimensional sorting (e.g. on both health and risk-aversion) might sometimes prevent or worsen adverse selection effects. Our basic 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 insurers generally have limited information, they tend to include only a limited number of potential objective risk predictors. Our contribution is that we include a much wider range, including information (e.g. lifestyle) that these studies would consider private. 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 self-perceived health. Or, they use very specific information such as the self-assessed risk to enter a nursing home in the coming 5 years (Finkelstein and McGarry, 2006). Instead, we deduce private information from a very broad subjective measure of general health.

Finally, there is an extensive literature on the quality of health insurance choices (see e.g., Bhargava et al. (2017)) and deductible choice. There are a number of papers focusing specifically 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 are more likely to choose the 'wrong' deductible. 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, which unlike Handel et al. (2023) does not take ex-ante averion against risk into account.

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 fact that until age 18 there is no (voluntary) deductible, they show that individuals with low ex-ante risk are more likely to select into the high deductible. They argue that it is mostly this selection effect and not the reduction in moral hazard that drives lower health expenditures among the high deductible group. Alessie et al. (2020) use a similar strategy as Finkelstein and McGarry (2006) and, using a small panel, conclude that private health information is predictive for costs and correlates with uptake of the high deductible. They also find that conditional on health, individuals with the high deductible use less health care, which they, contrary to the findings by Remmerswaal et al. (2023), interpret as evidence of moral hazard driving a part of the cost difference between individuals with the low- and the high deductible. In both studies, an implicit necessary link between individuals' risk of high health care costs and their choice is that they should know about their own health. We make this link explicit and disentangle individuals' knowledge about their health from their ability to use this knowledge when choosing their deductible level.

2 Institutional Setting and Optimal Insurance Choice

2.1 The Voluntary Health Insurance Deductible

In the Netherlands, every adult must take out a compulsory basic health insurance. While there is competition between insurers, the basic package of covered healthcare services is defined by the government. Health insurers cannot engage in risk-rating and there is a detailed risk-equalization scheme compensating insurers. All insurance contracts include a compulsory deductible of currently €385 per year, but individuals can choose to increase this in steps of €100 by up to €500 resulting in a maximum deductible of €885 (the compulsory €385 and the voluntary €500) in return for a premium discount. Insurance and deductible choice happens at the start of each calendar year. The average annual premium for an insurance package with the low deductible is around €1200 in our sample. Although different intermediary levels exist, most individuals choosing a higher deductible go for the maximum increase by €500 to €885 in total. Around 9% of all insured opts for a voluntary deductible. Of these the majority, more than 70% percent of individuals with a deductible, choose the maximum of €500 (Handel et al., 2023). The premium discount individuals receive in return for the high deductible depends on the insurer, but for the €500 increase the average is around €250 and the variation across insurers is small with most of the discounts ranging from €200 to €300 (Handel et al., 2023). Not all healthcare costs incurred are subject to the copayment components, but most costs are insurable with only certain parts of general practitioner care, all maternal care and all nursing care being always insured, not subject to any copayment when such care is consumed.

2.2 Optimal Choice in the Absence of Risk Aversion

We focus on the choice between the low deductible (€0) and the highest voluntary deductible (€500). In a standard expected utility framework, the choice an individual has to make 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 fixed premium) outweighs the expected utility with the high deductible (a lower fixed premium, but a higher risk of out of pocket costs). If individuals are risk-neutral, the optimal choice only depends on whether the expected payments (premium plus out of pocket costs) under the high deductible are lower than under the low deductible. This in turn only depends on the premium discount and the expected additional 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 will exceed €885. This means that individuals who choose the high deductible end up either having zero additional costs if they stay under the limit of the low deductible or face the maximum possible €500 additional costs if they exceed that limit. Expected additional costs under the high deductible are then a simple function of the probability that an individual i will stay within the low-deductible limit: $\mathrm{E}(c_t) = \mathrm{P}(c_i > \tau) * 500$. Because the premium discount received (€250) is half the level of the additional 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 straight-

forward decision 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 an individual's optimal choice and compare this to their actual choice. In case the actual choice deviates from the optimal, we can say individuals make a mistake (in the sense that their choice does not maximize their utility). However, this interpretation crucially depends on whether an individual's welfare is only driven by the expected costs and not by other things. The most important of these other factors are risk preferences.

2.3 Optimal Choice in the Presence of Risk Aversion

We can generalize the decision problem to allow for risk-aversion by introducing a utility function $U(.)$ for consumption. 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 now 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 it is optimal to choose the low deductible 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. For instance, 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 and interpret differences between the actual choice and the optimal choice under risk-neutrality as sub-optimal from the individual insured's perspective.

3 Data

3.1 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 covering various topics from healthcare use and costs, declared incomes from income tax data and labour market participation alongside demographic information like date of birth, registered gender and date of death.

Using administrative health records we construct a range of measures capturing individual-level health outcomes. Based on data from the national hospital admission register (LBZ), covering the universe of Dutch hospital admissions, 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 main chapters of the International Shortlist for Hospital Morbidity (ISHMT). Using data on annual prescription medication use by Anatomical Therapeutic Chemical (ATC) classification 4-digit level codes we apply the encoding developed by Huber et al. (2013) and refined by Danesh et al. (2024) to infer the presence of 21 different chronic conditions based on observed prescription medication use. Using electronic health claims data we observe total healthcare costs and costs broken down by the cost-categories covering GP care in total and for consultations only, hospital care, pharmacy-related 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 healthcare costs counting towards the copayment threshold. Using data on the universe of household-level and individual income records as reported to the tax authority we create information on relative income rank (deciles, quintiles and terciles) for 5-year age-gender specific groups based on per-capita household incomes of the entire population. Lastly, we observe the chosen deductible levels for the entire Dutch population between 2011 and 2021.

3.2 Survey Data

Our administrative data can be linked to various survey datasets using pseudo-anonymized unique individual-level identifiers and we make use of two of such 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 a large-scale, nationally representative survey of the general adult population in the Netherlands (aged 18+) aimed at measuring population health and organized by Statistics Netherlands. 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. The surveys are conducted between September 1st and December 31st 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 roughly 1,19 million unique individuals identifiable in administrative records. While the observed Health Monitor samples over-represent older individuals it is 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 costs above the copayment limit for 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 (normal, overweight, obese). Further the Health Monitor includes information on current health by eliciting 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 it includes the Kessler 10-item depression and anxiety questionnaire (Kessler et al., 1994), a clinically validated screening questionnaire for depression and anxiety related disorders allowing us to observe the raw score and the corresponding risk-categorization. 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 very broad measure of individuals' perceptions of their own health status 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 (Wuppermann, 2017), in particular when combined with other health information.

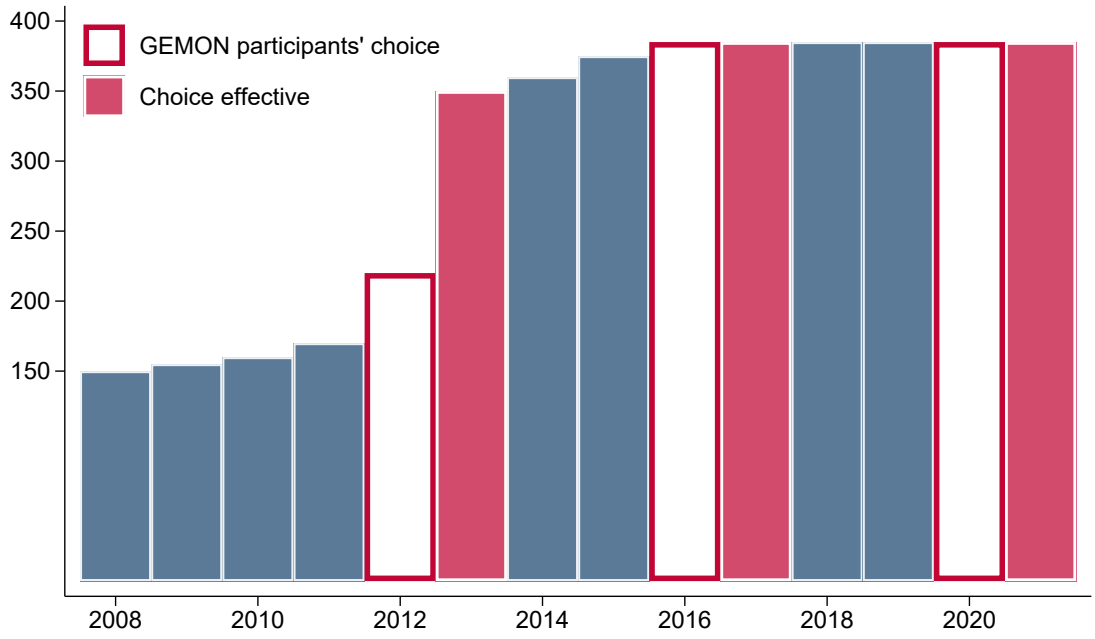
The LISS panel is a nationally representative household panel survey that follows individuals over time and covers similar variables as those included in the Dutch Health monitor alongside one-off survey modules covering other behavioural determinants of insurance choice such as risk aversion or subjective survival expectations. We make use of this data only in supplementary analyses to further explore empirical relationships identified based on the large scale Health Monitor surveys.

[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. Most responses were submitted online with in-person or phone-based interviews making 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.

3.3 Deductible Choice Timing and Analysis Samples

Figure 1 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 three Health Monitor surveys. As outlined by Handel et al. (2023) 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 actively advertise their products and send reminders to their customers. The choice at the end of the year is effective for the subsequent year with the deductible being applicable to the 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 in which case the initial choice is defaulted to. 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 1: Compulsory Deductible Levels and Health Monitor Timing, 2008-2021



Source: Own calculations based on publicly available data.

Using the administrative and survey data we create two datasets. Firstly, a dataset of the entire Dutch population aged between 21 and 80 on January 1st 2016 and surviving at least until January 1st 2017, covering 11,547,759 unique individuals for which we can observe voluntary deductible levels in 2016[2] and administrative data for the two preceding years (2015 and 2014). Secondly, we create a pooled Health Monitor sample combining survey participants from 2012, 2016, and 2020 if they were aged between 21 and 80 on January 1st of the survey-year and

[2]Focusing on 2016 is for now simply a pragmatic choice and in principle we can create similar datasets also for the other years of the Health Monitor survey rounds.

survived at least until January 1st in the subsequent year. Furthermore we condition on the availability of the same administrative records as for our 2016 general population dataset and the availability of all data from the Health Monitor survey as outlined above. This yields an analysis sample of 837,697 unique Health Monitor respondents.

4 Empirical Strategy

Our primary research question is whether private health information allows individuals to make optimal decisions in the uptake of the voluntary deductible, thereby improving choice quality from the insured's perspective but potentially resulting in adverse selection. To explore this we divide our empirical strategy into three steps. Firstly we create a measure of health risk based on health 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 to serve as a benchmark against which the added informational value of private information dimension in making optimal choices can be compared (Section 4.1). Following Handel et al. (2023) we use of cross-section of the entire adult Dutch population to calibrate age-gender specific health risk prediction models. In a second step we use the Health Monitor survey data to quantify the degree of private health information individuals have based on data on self-assessed health, lifestyle and health behaviors and other privately observed health information in combination with the created health risk model (Section 4.2). Thirdly, we combine information on private and mutual health information and the predicted and actual realisation of health outcomes and deductible choices to explore whether private health information allows individuals to make individually optimal insurance choices (Section 4.3).

4.1 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 2.2, individuals' optimal deductible choice depends (solely) on the risk of health care costs in year t being above or below the deductible threshold τ. From here onward we focus on the case of being an individual with costs below the copayment threshold, specifically with costs below €375[3] in all empirical analyses. 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. The same optimal choice-rules as outline in Section 2.2 apply irrespective of this choice.

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 a rich set of candidate predictors mutually observable by the insurer and the insured. Consider the vector of mutually observable health information $X_{M,i}$ in the years $t-1$ and $t-2$. In our context it contains information on past healthcare costs by category, hospital care use by ISHMT group and the presence of chronic conditions identified using medication use data. We use this data to estimate a linear probability model:

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

[3] As shown in Figure 1 the compulsory deductible has varied over time but remained stable at €385 since 2016. We chose €375 as it is the mean value across our observation period but plan to make use of year-specific copayment thresholds in the future. Given the small differences in the compulsory deductible for most of our observation periods this choice is likely of small importance.

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

The resulting age-gender specific prediction models are then applied to the Health Monitor 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 $\hat{\rho}_M = \mathrm{P}(c_i \leq \widehat{\tau} | X_{M,i})$ in the subsequent years, 2013, 2017, and 2021 respectively.

4.2 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 individuals have *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 but known to the insured, $X_{P,i}$ in the form of information on health behaviours, functional limitations, and mental health measures elicited in the Health Monitor surveys.

The creation of our measure of private health information is relatively 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 the following linear regression 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 privately observable 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 both our measure of health risk using mutual health information ($\hat{\rho}_{M,i}$) and the different dimensions of privately observed health information ($X_{P,i}$) covering information on information on health behaviours, functional limitations, and mental health. 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 between health dimensions in $X_{P,i}$, age and gender. In all specifications and those outlined from here on 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[4]. Ultimately, however, these cut-off values are chosen somewhat arbitrarily. If individuals respond with subjective health assessments above or below what can be explained by mutually observable $X_{M,i}$ and privately known $X_{P,i}$ 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 risk of health care costs.

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 once we condition on other health information. 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 health information and other privately observed health 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).

4.3 Private Information and Insurance Choice

4.3.1 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 (see Section 4.1) with our measures of private health information (see Section 4.2). Combining data on health risk predicitons 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 another set of age-gender specific linear prediction models to predict individuals likelihood of having costs below the copayment threshold:

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

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

Note that we only include dummies for membership of the subjective health class $S_i = 1$, individuals indicating higher-than-expected self-assessed health, and $S_i = 3$, individuals indicating lower-than-expected self assessed health, using the neutral group ($S_i = 2$) as the reference group. We further considered two alternative specifications to aggregate 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 ($X_{M,i}$) and privately observed health characteristics ($X_{P,i}$). 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 ($X_{P,i}$) to vary between the subjective health classes. Of these the fully interacted model resulted in the best prediction performance, however only marginally better than the inclusion of the dummies. All our results are consistent across these specifications and in the results section we present only the results from Equation 5.

Estimating Equation 5 allows us to explore two things. The updated individual-level health risk prediciton $\hat{\rho}_{MPS,i} = \mathrm{P}(c_i \leq \tau | \hat{\rho}_{M,i}, X_{P,i}, S_i)$, and the direct effect of membership in a specific subjective health class S_i on the likelihood of having costs below the copayment threshold. To also consider the importance of private health information on the uptake of the voluntary deductible we follow a similar estimation strategy be estimating the following linear probability model:

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

Where $d_i = 1$ indicates that an individual has chosen a deductible ≥ 0. If subjective health has predictive power for both the likelihood of having costs below the copayment threshold (Equation 5) and for the uptake of the deductible (Equation 6) this is sufficient to establish the presence of private information and resulting selection effects as we can observe both health risk and the chosen insurance coverage (Finkelstein and Poterba, 2006). If better ($S_i = 1$) or worse ($S_i = 3$) than expected self-assessed health has significant predictive power indicating 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 a voluntary deductible we would observe adverse selection. In the context of our decision problem, healthier individuals opt for lower coverage while less healthy individuals opt for high coverage where the marginal cost for care consumption is reached at the lowest possible copayment threshold. If health risk and coverage choice would show an opposing correlation we would observe advantageous selection with healthier individuals opting for higher coverage despite not needing it and vice-versa for unhealthy individuals.

4.3.2 Private Information and Choice Quality

While our empirical approach outlined in Section 4.3.1 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

from the individual's perspective. Even if the correlation is in the right direction, individuals might under- or overreact in their deductible choice. In the fourth step of our empirical analysis we therefore consider the interaction between our measures of health risk and deductible uptake consistent with individual-level health risk *after taking into account private health information* individuals hold.

To achieve this we make use of a relevant feature of our particular decision task. As outlined in Section 2, individuals should choose the high deductible if the probability that their healthcare expenditures remain below the copayment threshold is higher than 0.5 and the compulsory deductible only if this probability is lower than 0.5. We can use Equation (5) to make individual predictions based on all available information $\hat{\rho}_{MPS,i} = \mathrm{P}(c_i \leq \tau | \hat{\rho}_{M,i}, X_{P,i}, S_i)$ together with observed deductible choices to create a choice quality measure:

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

If $Q_i = 0$ an individual's choice is optimal, if $Q_i = -1$ an individual is overinsured (low deductible, while $\widehat{\rho}_{MPS,i} > 0.5$), and if $Q_i = 1$ an individual is underinsured (high deductible, while $\widehat{\rho}_{MPS,i} > 0.5$). To directly 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 overinsured or underinsured, so $Q_i \neq 0$. This focus on the *ex-ante* optimal choice individuals make allows us to also consider the heterogeneity of choice quality *within* subjective health classes and across different health risk-levels of which we have an empirical estimate $\widehat{\rho}_{MPS,i}$. Even among those who report better health than predicted, there are some with a low probability of having costs below the copayment threshold ($\widehat{\rho}_{MPS,i} < 0.5$). Similarly, among those who report worse health than predicted there are individuals with a high likelihood of low costs ($\widehat{\rho}_{MPS,i} > 0.5$). This allows us to compare the choice quality of individuals with *similar* health risk levels but in a different subjective health classes.

Focusing on the *ex-ante* choice quality explicitly allows the quality of our health risk predictions 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* actual 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 $c_i \leq \tau$.

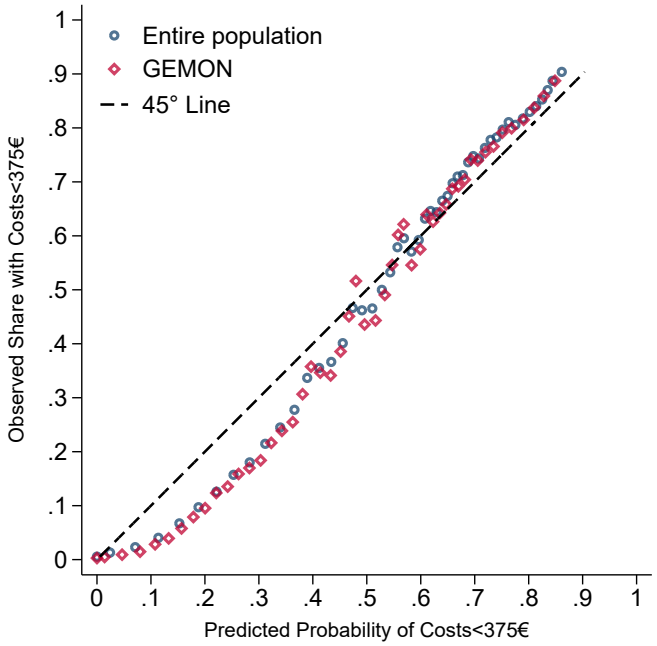
5 Results

Our results are presented in line with the different steps of our empirical analyses as outlined in the methodology section. First we present evidence on the performance of our general health risk prediction model used to estimate individual-level health risk using health information observed by both the insurer and insured (see 4.1). Secondly we present results from our empirical models aimed at quantifying individual-level private information based on the Health Monitor Survey Data (see 4.2) and whether the inclusion of this information adds additional relevant explanatory power with respect to health risk and deductible uptake beyond the mutual health information (see 4.3.1). We then present results on the quality of choices taking into account this private information (see 4.3.2). Lastly we present evidence from additional supplementary analyses.

5.1 Health Risk Prediction Model Performance

Figure 2 presents evidence on the performance of our LASSO-based age-gender-specific cost-prediction models (see Section 4.1) 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 actually 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 on average yield predictions in which average linear probabilities are equal to the actually observed share of actually realized healthcare costs.

Figure 2: 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 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). Predicted linear 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.

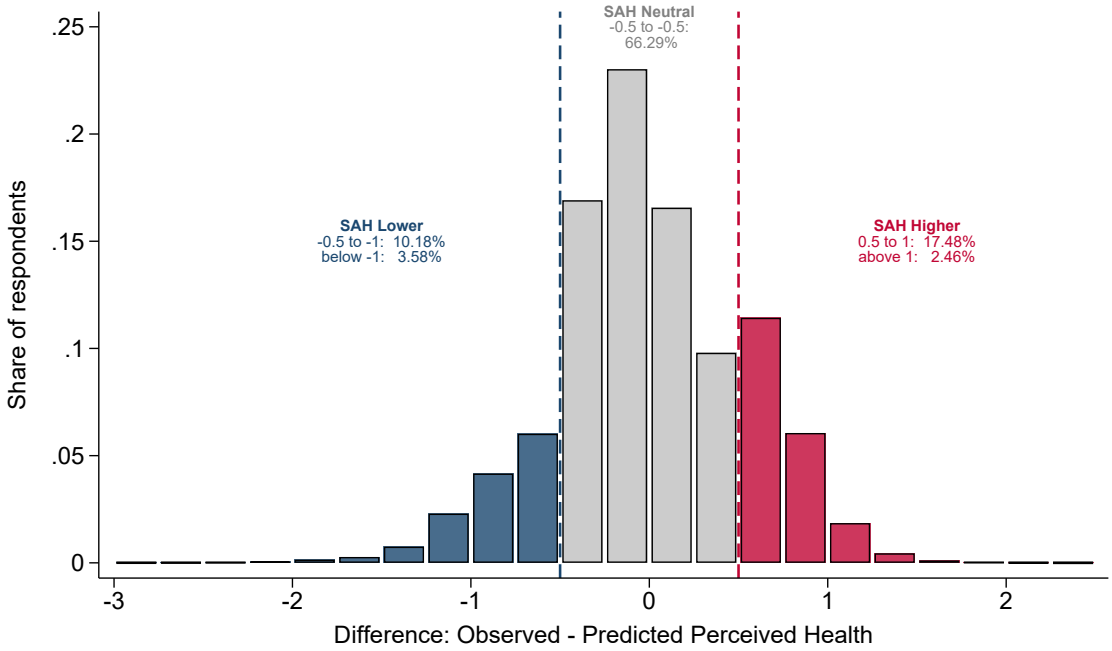
Overall our prediction model performs well. However, among those with higher health risk (having a lower likelihood for costs below the threshold) our model tends to over-estimate while among those with lower health risks it slightly under-estimates with predicted probabilities being persistently below observed realizations. 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 actual deductible choices (see Figure A1.2) we can observe that for those actually taking the voluntary deductible the observed share of individuals with low costs is consistently above the predicted shares. This provides further evidence on the potential importance of private information in explaining choices and choice quality.

5.2 Private Health Information, Health Risk and Deductible Choices

To explore the extend of private health information individuals have and whether this information offer additional explanatory power over mutually observable health information $X_{M,i}$ we first estimate a measure of privately held health information as outlined in Section 4.2. Figure 3 depicts the residuals (prediction errors) from the regression of self-assessed health on the linear probability of exceeding the copayment threshold $\hat{\rho_M}$ and the private health information $X_{P,i}$ observed only 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 first,

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, and those in between as equal as predicted or neutral. The broad majority of individuals' are classified as SAH-neutral with 68.29% of the sample falling into this category, 13.76% of individuals are classified as SAH-lower and 19.94% as SAH-higher.

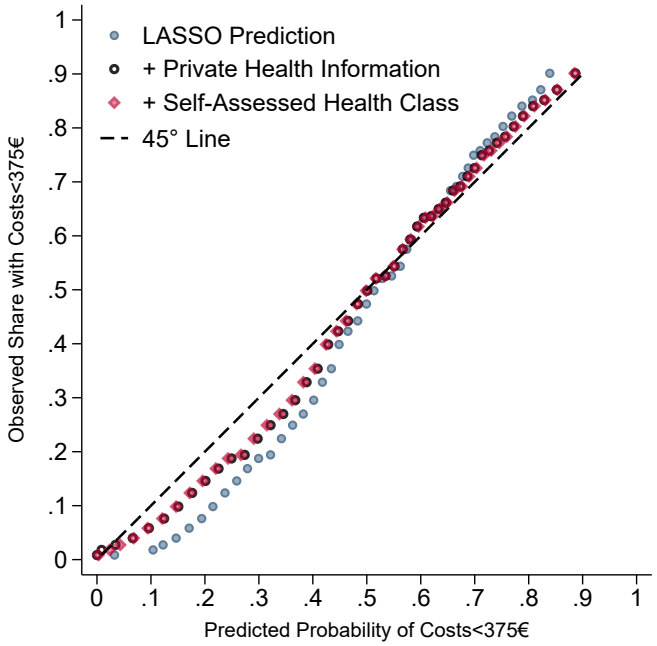
Figure 3: Self-Assessed Health Residual Distribution and Subjective Health Classes



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

Figure 4 provides an overview on the performance of our health risk prediction model when including private health information. As Figure 2 it depicts on the y-axis the observed share of individuals with healthcare costs below the copayment threshold against the predicted likelihood on the x-axis. It contrasts 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 that is not mutually observed improves the performance of our cost-prediction model considerably while the inclusion of SAH-class membership provides an additional but smaller performance improvement. This provides evidence for the general importance of private information in predicting health risk.

Figure 4: Health Risk Prediction Model Performance with Private Health Information



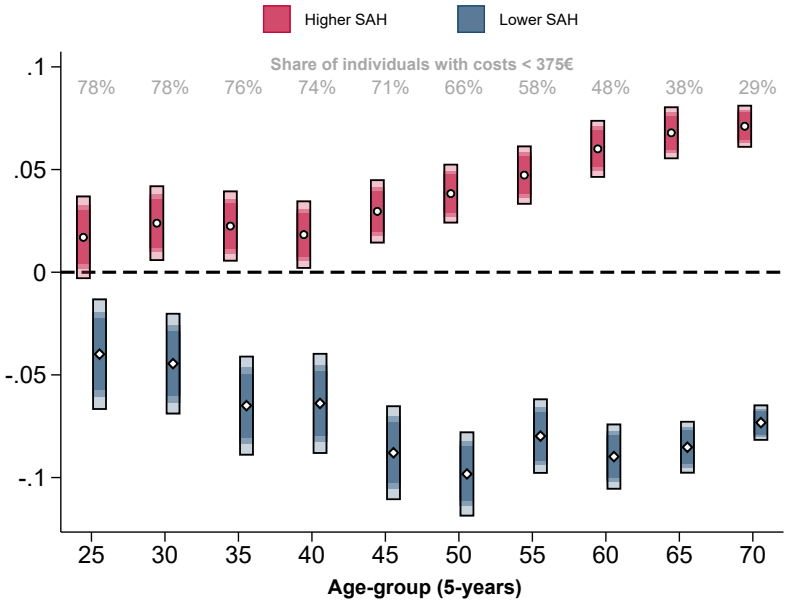
Source: Own calculations based on data from Statistics Netherlands. *Note:* Predicted linear probabilities come from a prediction model including self-assessed health, the linear probability from the general population health risk model and health information contained in the Health Monitor surveys (health behaviours and physical/mental health) alongside single-year-of-age and survey-year dummies. All regressions are done by 5-year-age-gender groups for the pooled sample of Health Monitor participants with N=837,697.

To explore the role of private health information in our health risk predictions and the uptake of the voluntary deductible across different groups of individuals we can also plot the coefficients from our regressions directly. Figure 5 plots the coefficients for subjective health class membership from Equation 5 membership for males (panel a) and females (panel b) by age-group to explore their predictive power for health risk. The plots also depict the average share of individuals with costs below the copayment threshold by age-gender group as a benchmark for the interpretation of coefficient sizes. 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 contained in S_i is most relevant for older individuals.

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

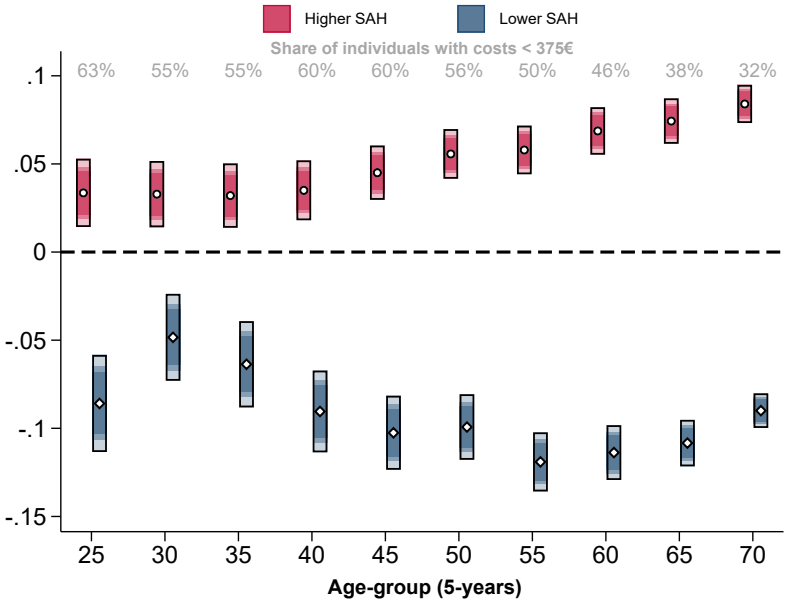
Figure 5: Subjective Health Classes and Health Risk by Age

(a) Males



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

(b) Females

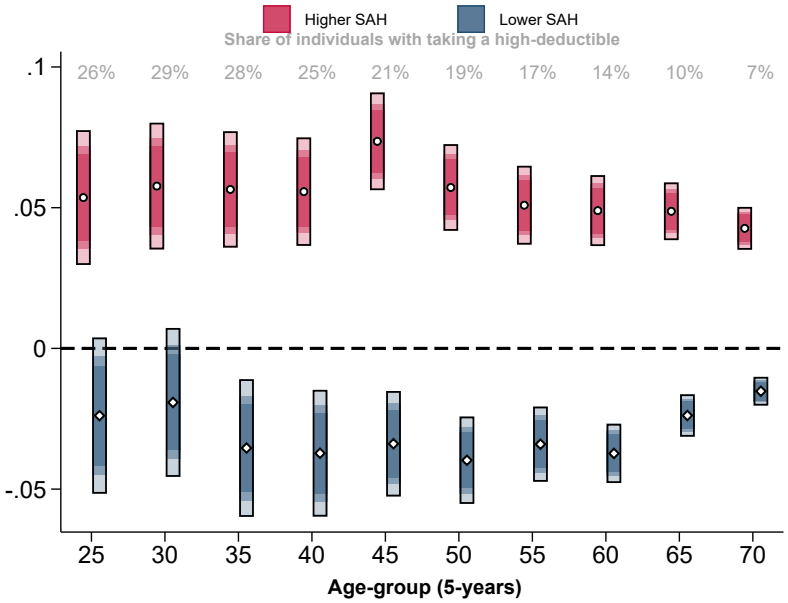


Note: Shaded bars indicate 90/95/99% confidence intervals.

Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the coefficients for subjective health class membership from a regression of low-cost-status in the year after the survey participation including subjective health-class dummies, the linear probability from the general population health risk model and health information contained in the Health Monitor surveys (health behaviours and physical/mental health) alongside single-year-of-age and survey-year dummies. All regressions are done by 5-year-age-gender groups for the pooled sample of Health Monitor participants with N=837,697.

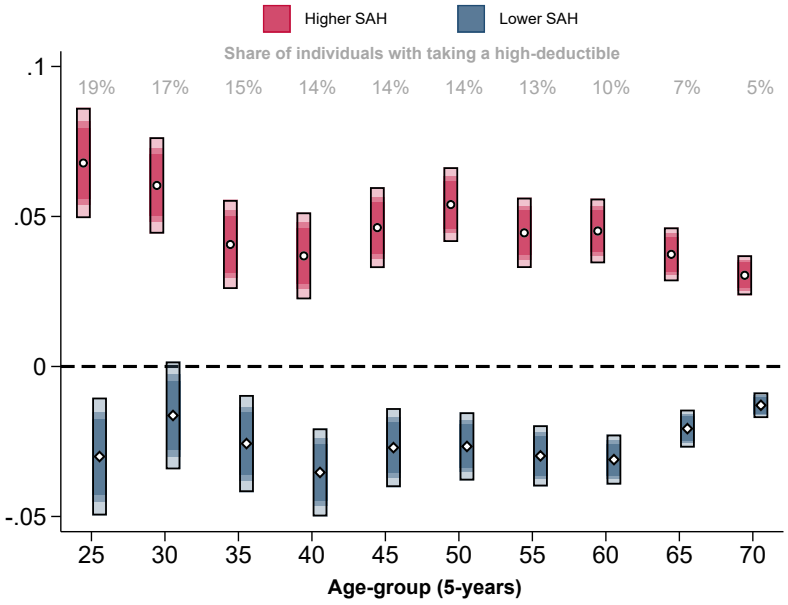
Figure 6: Subjective Health Classes and Deductible Uptake by Age

(a) Males



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

(b) Females



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

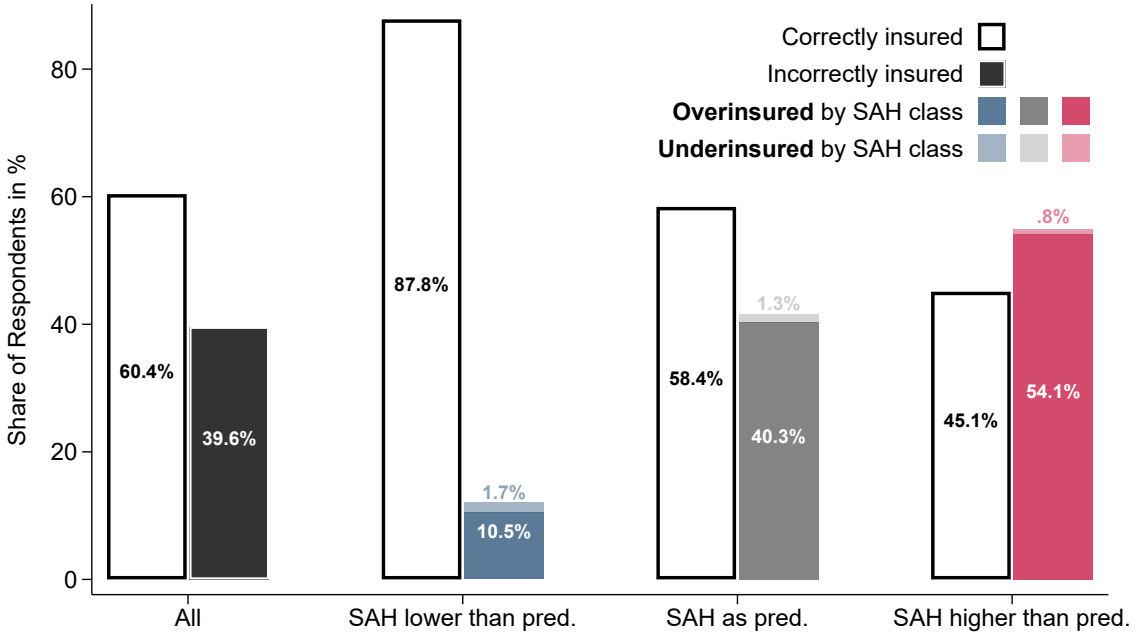
Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the coefficients for subjective health class membership from a regression of deductible uptake in the year after the survey participation including subjective health-class dummies, the linear probability from the general population health risk model and health information contained in the Health Monitor surveys (health behaviours and physical/mental health) alongside single-year-of-age and survey-year dummies. All regressions are done by 5-year-age-gender groups for the pooled sample of Health Monitor participants with N=837,697.

5.3 Private Information and Choice Quality

Our results thus far have established 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 (Figure 5) and deductible uptake (Figure 6) suggest there is evidence for the effect on actual insurance choice to lag behind the actual predictive power in terms of health risks, especially for older individuals. To what extent does this also lead to better economic decision making and a higher choice quality?

Figure 7 shows our main results regarding choice quality by providing a simple descriptive view of choice quality for the entire sample and by subjective health class. Overall, the majority of individuals chose a deductible level in-line with their predicted healt hrisk, with 60.4% being correctly insured. However, a large minority of 39.6% are incorrectly insured of which the overwhelming majority, 38.4% percent of the population, are overinsured compared to a small population share of 1.2% being underinsured. Breaking down these shares by subjective health classes reveals that there is a large difference. While among individuals with lower than expected self-assessed health 87.8% chose the right deductible level in line with their health risk this share is only 58.4% for those individuals with a neutral subjective health class but a minority among those with a higher subjective health class where only 45.1% chose the right deductible. These results highlight the complex interaction between knowing about your health and optimally using this information. There seems to be, as documented (Handel et al., 2023), a tendency among the Dutch population to avoid the high deductible.

Figure 7: Choice Quality by Subjective Health Class



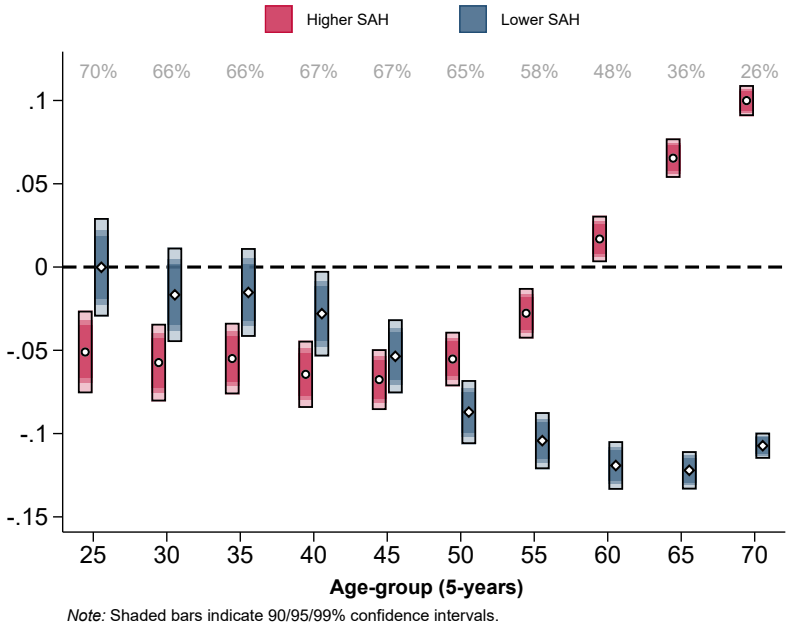
Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the share of respondents by *ex-ante* based choice quality for the full sample (N=837,697) split into correctly and incorrectly insured, and each of the subjective health classes SAH higher (N=167,036), SAH neutral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

Turning to the direct impact of self-assessed health on the quality of choices across we next present our results from estimating Equation 7 for the general case of being incorrectly insured (either over or under-insured). Figure 8 depicts the subjective health class coefficients for our choice quality measure using the *ex-ante* perspective. As can be seen we observe a clear age-specific pattern of results which is present for both male and female respondents but 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 lag behind the predictive power individuals private health information contains. The results on our measure of choice quality confirm however that the implied pattern of average choice quality also shown in Figure 7 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, as can be seen in Figure 9. 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 and across levels of health risk. To do so Figure 10 plots the share of individuals taking the high deductible on the y-axis against the predicted linear 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 risks 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 in general they do so insufficiently, in particular among 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.

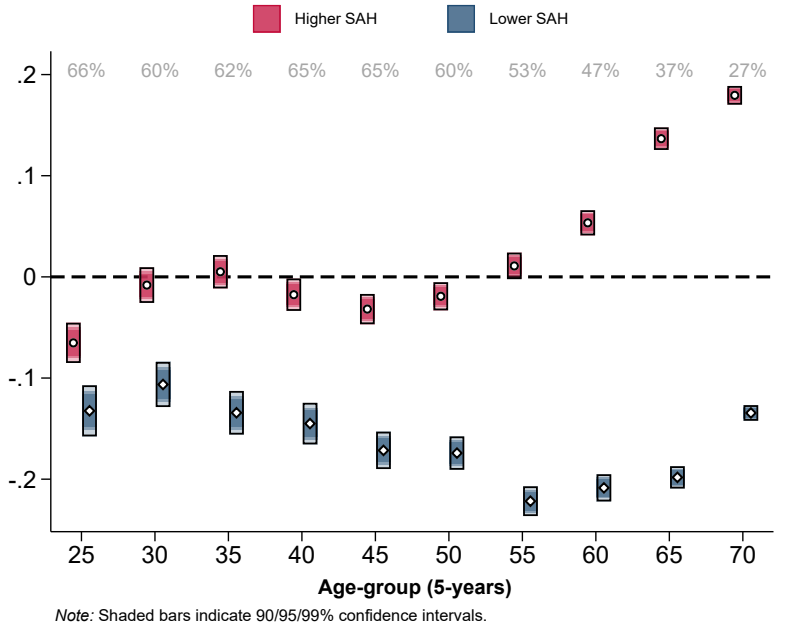
Figure 8: Subjective Health Class and *Ex-ante* Insurance-Errors by Age

(a) Baseline Predictions



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

(b) Including Private Information

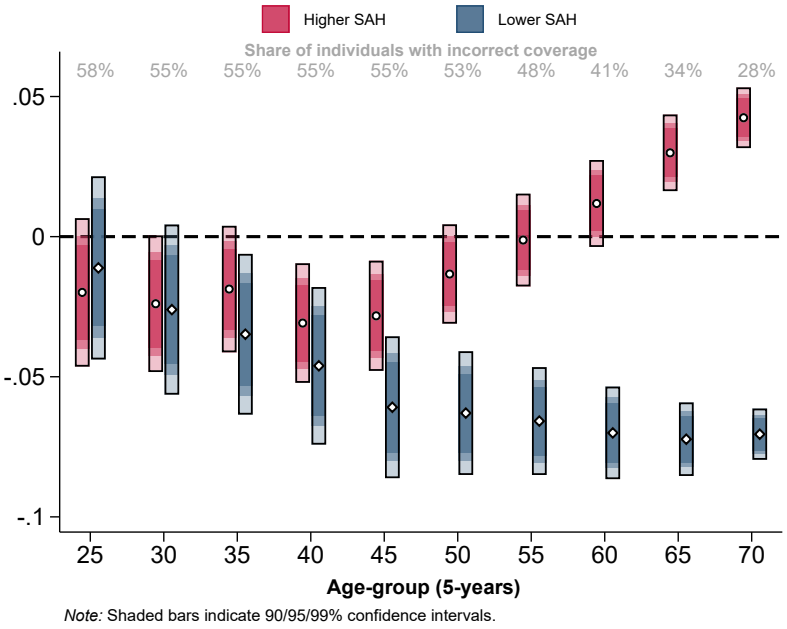


Note: Shaded bars indicate 90/95/99% confidence intervals.

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

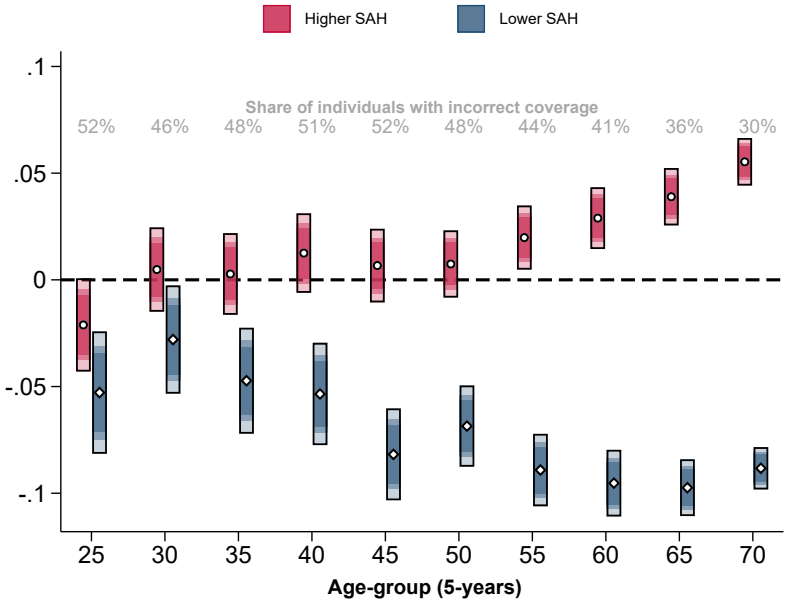
Figure 9: Subjective Health Class and *Ex-post* Insurance-Errors by Age

(a) Baseline Predictions



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

(b) Including Private Information

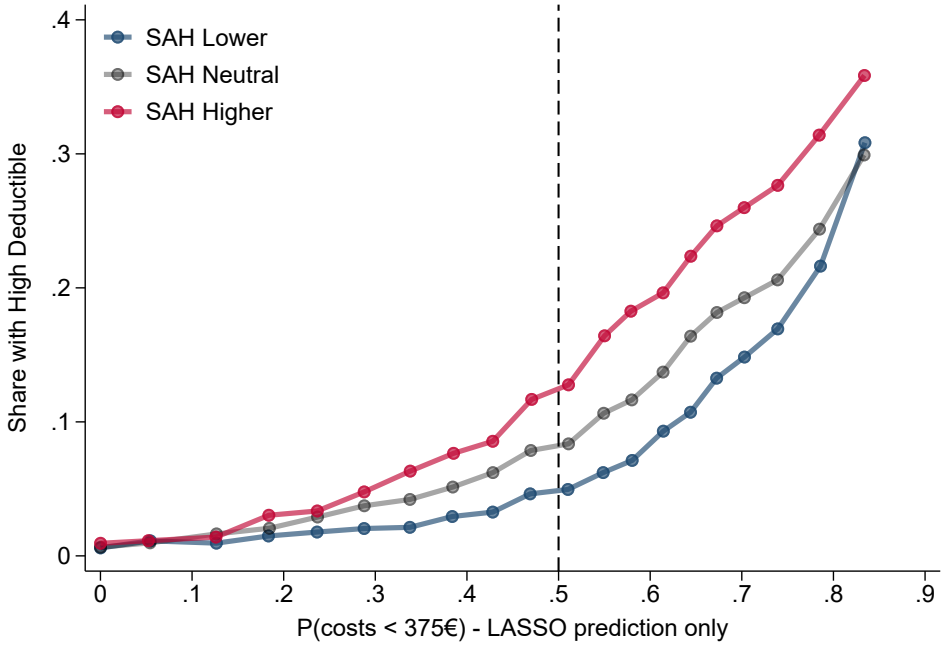


Note: Shaded bars indicate 90/95/99% confidence intervals.

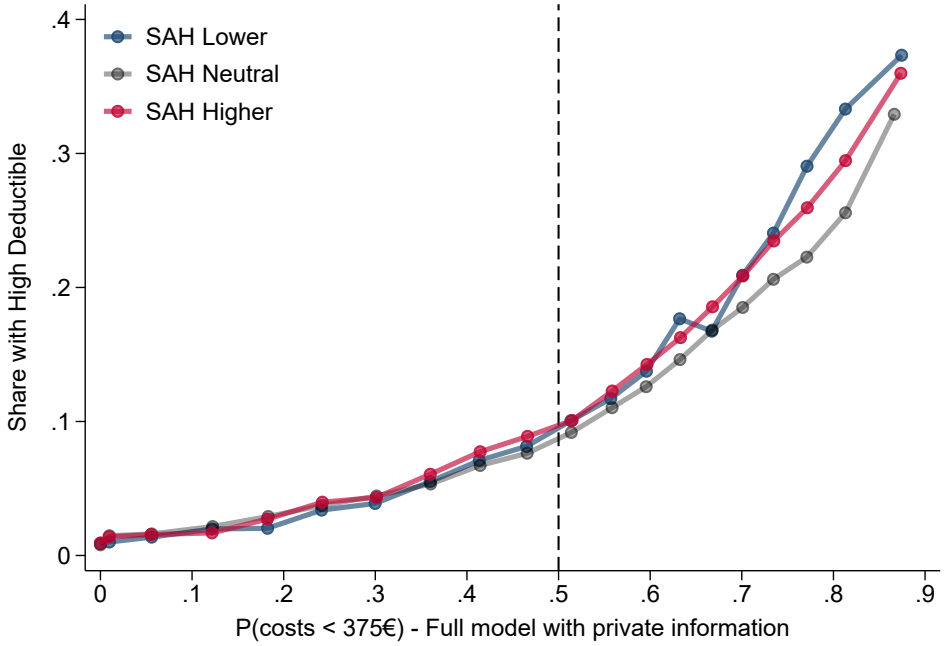
Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the coefficients for subjective health class membership from a regression of the *ex-post* choice quality measure indicating an incorrect deductible choice in the year after the survey participation including subjective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

Figure 10: 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.

5.4 Mechanisms and Additional Analyses

5.4.1 Socio-Economic Status and Education

As shown by Handel et al. (2023) the choice quality of deductible choices in the Netherlands is heavily impacted by an education and socio-economic gradient. They find that individuals with higher incomes and better education levels while still overinsure on average 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. We therefore consider the role of both income and education. Firstly we can run our entire analysis while accounting for education levels and income status. To explore how the choice quality patterns we have shown so far play out once divided by education and income levels. Figure A1.3 in the Appendix replicates Figure 10 but dividing the sample both by subjective health classes and also 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. The latter, however, we consider in the next section with little evidence for this to be a decisive factor.

5.4.2 Risk Aversion

Another potentially salient driver of deductible choices and the degree of overinsurance we observe is risk aversion. As outlined given the small monetary amount involved in the decision taks 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 indeed 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 simply summing up the number of sure choices an individual made (Alessie et al., 2020). The risk aversion measure derived from this ranges from 0 to 5, with higher numbers indicating more risk aversion. 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 an effective sample of around 4000 unique individuals and 15660 individual-year observations. As outlined in the data section the LISS survey cover all variables necessary to replicate the analyses done with our large

Health Monitor sample with the smaller LISS 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.

6 Discussion & Conclusion

In this paper we provide evidence on the role of private health information in determining the quality of insurance choices. Combining large-scale survey data with administrative records we analyzed the quality of health insurance deductible choices in the Netherlands. Our primary goal was to disentangle the welfare implications of private health information for insurance choice by differentiating between mutually observable and private health information, whether this information improves health risk predictions and results in individually optimal insurance choices.

Our empirical results provide three key insights into the role of private health information in determining choice quality. Firstly, we show that individual's hold considerable private health information 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, in our context their deductible level, but do so to a far lesser degree than what would be optimal. Thirdly, as a result of this overall choices are far from optimal from the insured individual's perspective, despite a considerable degree of private information being available. We find that individuals with private information indicating a low risk of future healthcare needs tend to be over-insured. Interestingly, individuals who report a lower level of health than predicted but still have a low risk of high costs more often choose a high deductible than those who report a higher than predicted level of health with the same risk.

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 in a way that is optimal for the individual. Individuals can make mistakes and over- or under-respond to the information they have. This is exactly what we find for individuals with private information that they are at low risk of high health care costs: although they are more likely to choose a high deductible than other groups, they are still largely over-insured.

Our results also have two implications for the design of the deductible. First, Remmerswaal et al. (2023) have raised the concerns that individuals with a low risk of high health costs self-select into choosing the voluntary high deductible. And thus, the high deductible might mainly function as a way for low-risk individuals to partly opt-out of the mandatory solidarity between healthy and unhealthy within the Dutch health insurance system rather than improving efficiency through lower moral hazard. However, the risk of positive self-selection into high deductibles seems surprisingly limited in practice and thus the welfare loss may be limited

too. We find that most individuals, even those who have private information that they are at low risk, opt for the low deductible. While in other contexts, for example the reaction to ambiguous health information, the inability of individuals to correctly act upon the information they receive can results in costly consumption of in-appropriate low-value care (Hoagland, 2021) we show that similar frictions can also result in individually sub-optimal decision patterns that are beneficial for the system as a whole.

Secondly, Handel et al. (2023) have documented socioeconomic inequalities in optimal deductible choice. 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: ability to use private information or risk aversion. We do find some 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 individual risk predictions and condition on the same risk levels. However, overinsurance levels still remain very high conditional on having private information irrespective of education or income levels.

6.1 Planned Extensions

Our study also has limitations which guide our ongoing work and plans on future extensions. Firstly, the correct prediction of individual-level health risk is a central component of our empirical strategy. While our current prediction models perform very well there is room for improvement. In particular we are considering to fully replicate the ensemlble machine learning model used by Handel et al. (2023) while also introducing a bayesian updating based adjustment of our 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.

A more important set of limitations however relates to us currently only focusing on a few alternative choice determinants and barriers. While we find little evidence for risk aversion as the main driver of our results on overinsurance. We plan to extend these by also considering relative risk aversion. However, 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 we can observe, 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 also containt information on locus of controls, a behavioral determinants that has been linked to the demand for supplementary health insurance (Bonsang and Costa-Font, 2022) and has been found to moderate individual-level behavioural responses to health events (Schurer, 2017).

Lastly we plan to expand our empirical analysis by considering the role of information shocks. The rational for this is that ideally we would want to explore the role of deductible changes keeping private information constant to also consider the role the deductible itself plays in potentially encouraging or discouraging healthcare demand (see also Alessie et al. (2020)) and thereby allowing for moral hazard. One potential way to do so is to consider those individuals experienced short one-off healthcare costs shocks in the past (see e.g. Bonekamp and Wouterse (2023)), thereby shifting individuals likelihood to pick a certain deductible level while keeping private health information on health risk constant. Such an approach would also allow us to directly explore the welfare effects of incorrect deductible adjustments conditional on private health information directly. Repeated information from LISS but also the small subsample of roughly 67.000 unique individuals participating in consecutive Health Monitor surveys are particularily interesting in this context.

6.2 Conclusion

In summary our study shows that individuals possess diverse but relevant private health information. Even when accounting for behavioural risk factors or other health characteristics only observable to the individual and not the insurer in our study context simple measures of self-assessed health still contain 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 overinsurance. This suggests that while pirvate information might remaine 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 and thanks to this individuals tend to make individually suboptimal decisions by not using the available information they have to the fullest degree. Which choice barriers and frictions are driving these results and to what extend it can be replicated in other insurance and country contexts will provide a fruitful research area going forward.

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.

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.

Bonekamp, J. and Wouterse, B. (2023). Do different shocks in health matter for wealth? *Journal of Health Economics*, 87:102719.

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.

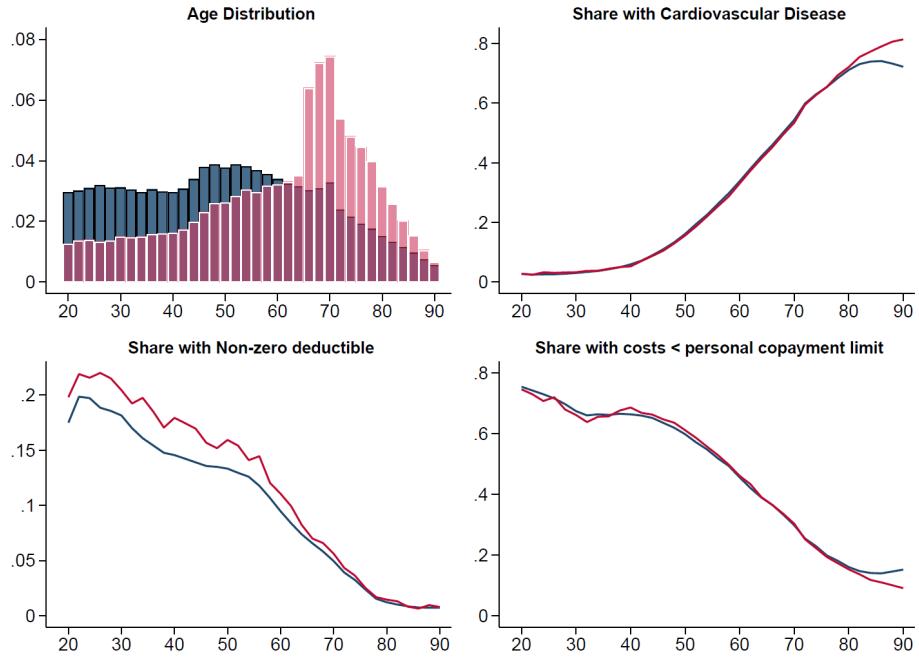
Van Winssen, K., Van Kleef, R., and Van de Ven, W. (2015). How profitable is a voluntary deductible in health insurance for the consumer? *Health Policy*, 119(5):688–695.

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.

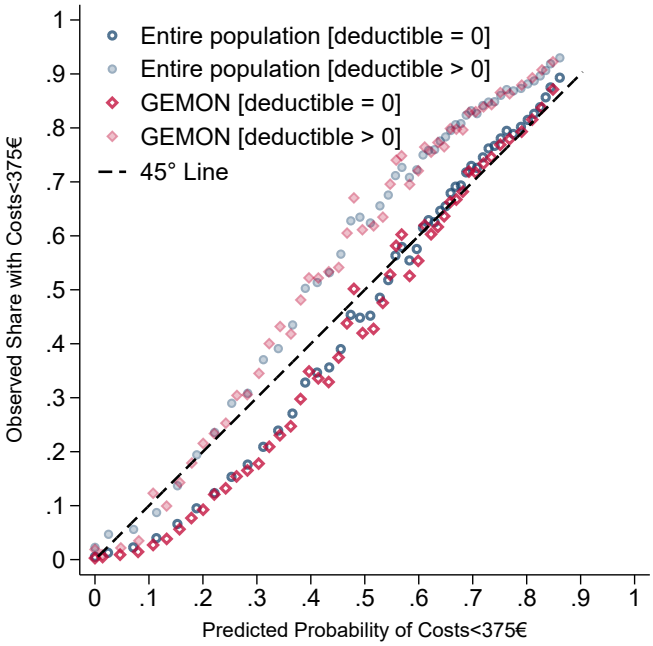
A1.1 Additional Tables and figures

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.

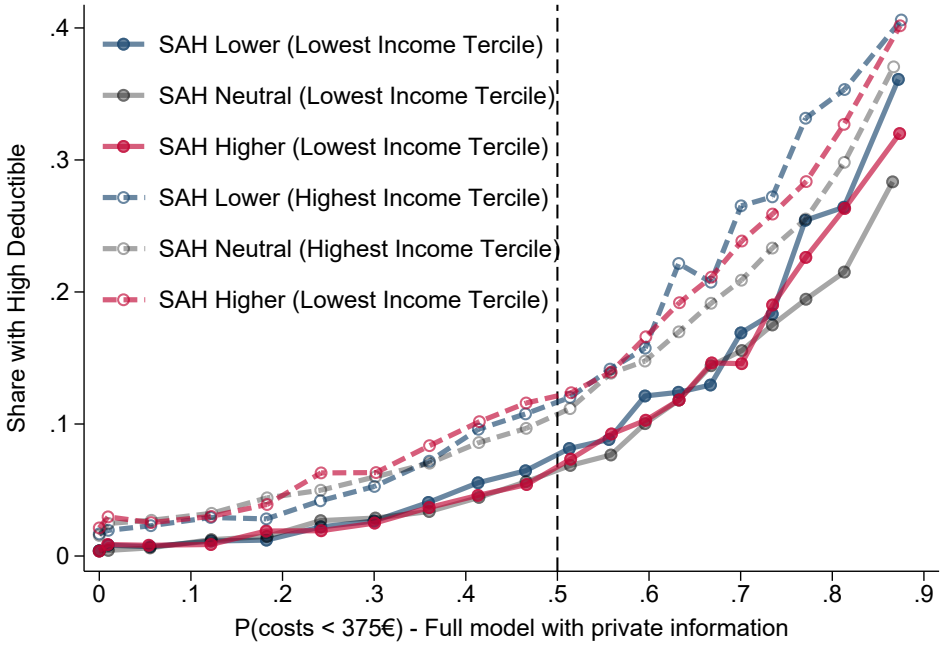
Figure A1.2: Health Risk Prediction Model Performance by Deductible Choice



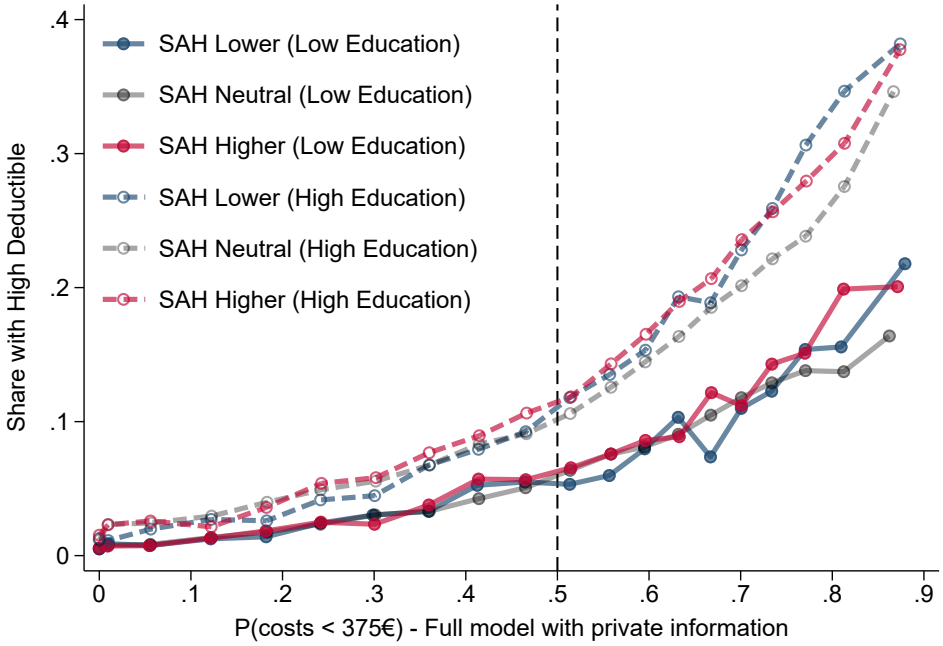
Source: Own calculations based on data from Statistics Netherlands. *Note:* The entire population sample contains all individuals aged between 20 and 80 years on January 1st 2016 (N=11,547,759) while the GEMON sample contains all unique participants of the Health Monitors 2012/2016/2020 aged between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