The Social Determinants of Choice Quality: Evidence from Health Insurance in the Netherlands*

Benjamin Handel, UC Berkeley

Jonathan Kolstad, UC Berkeley

Thomas Minten, London School of Economics

Johannes Spinnewijn, London School of Economics

October 11, 2019

Abstract

There are a range of industries and contexts where regulators directly impact the product choices that consumers have available. If consumers make choice errors, as much prior work documents, then market designers will generally want to weigh the improved match benefits of additional choices with the potential negative consequences of greater consumer misallocation. We study these issues in the Netherlands health insurance market, where the regulator specifies what deductible levels private insurers can offer consumers. Our analysis studies all consumers in the Netherlands using rich administrative data on health insurance choices, health utilization, education, income, wealth, and peers. We document stark differences in choice quality across these domains finding that (i) higher education levels (ii) more analytic degrees (iii) more analytic professions (iv) higher income and (v) higher wealth are all associated with better decisions, conditional on health risk and myriad other socio-demographic factors. We show that peer effects, within families and within firms, have meaningful impacts on consumer choice quality. We leverage these results to study the normative implications of several counterfactual scenarios including (i) optimal allocation of consumers to deductibles (ii) a high deductible mandate and (iii) a low deductible mandate. We assess the consumer surplus and equity implications of these policies and find that, while consumers leave most of the surplus from potential deductible matches on the table, the mandates we consider reduce consumer welfare. We show that the equity implications of these policies depend critically on (i) choice quality by socio-demographic group conditional on health risk and (ii) the correlation between health risk and socio-demographic characteristics.

*We thank the Central Bureau of Statistics of the Netherlands and especially Annemieke Redeman for help with the data. Chloé de Meulenaer and William Parker provided excellent research assistance. We gratefully acknowledge funding by ERC, ESRC and STICERD.

I Introduction

Consumer choice is a core aspect of the provision of many important regulated products. In many contexts, retirement investments (see, e.g., Hastings et al. (2013)), schooling (see, e.g., Neilsen (2017), electricity (see, e.g., Ito (2015)), and health insurance (see, e.g., Enthoven, Garber and Singer (2001)) regulators facilitate consumer choice, sometimes via private providers, under a set of restrictions governing what those choices are. A key objective is to set up choices that allow efficient heterogeneous matching without making the choice problem so complex that consumers frequently choose options that are sub-optimal for themselves.

While there have been many studies across these sectors investigating consumer choice quality, these studies typically look at a specific slice of the population and/or do not have much information on consumer socio-demographic characteristics. [1] When thinking about the tradeoff between greater product availability and choice complexity, regulators likely want to consider both population average outcomes but also outcomes for specific sub-classes of consumers. Specifically, policymakers generally are concerned with the redistributive implications (e.g. across income percentiles) of regulated product provision, as well as the efficiency implications. Without both (i) broad information on socio-demographic characteristics and (ii) a clear understanding of what constitutes a "good" or "bad" product allocation it is difficult to analyze the redistributive implications of regulatory policies that engage consumer choices.

In this paper we investigate the consumer surplus and equity implications of insurance provision in the Netherlands, where the insurance market is highly regulated but some significant degree of consumer choice remains. The Netherlands insurance market is a managed competition environment that is similar in spirit to, e.g., the Affordable Care Act, in the United States, though in Netherlands regulation of how insurance contracts can vary is more restrictive than in the United States. One key choice dimension in the Netherlands is the insurance contract deductible, which specifies the amount in each year a consumer must pay out-of-pocket before insurance payments kick in. Specifically, all insurance contracts have a baseline deductible (375 EUR in 2015) but consumers can elect higher deductibles in 100 EUR increments up to an additional 500 EUR (875 maximum total deductible in 2015). When consumers elect a higher incremental deductible, they get a premium rebate that is typically equal to about half of the incremental deductible amount.

We leverage granular comprehensive data on all consumers in the Netherlands ($N \approx 12$ million) to assess the consumer surplus and equity implications of Netherlands' deductible choice policy. We observe what kind of insurance each consumer elects, detailed individual-level health information, and detailed individual-level socio-demographic information including (i) income (ii) wealth and debt (iii) their educational background (iv) their profession (v) their firm (vi) their location and (vii) the insurance choices of their parents and relatives. The

[1] Chetty et al. (2014) is a notable exception in the domain of retirement savings.

depth and scope of our data allow for both precise assessments of consumer deductible choice quality and an investigation of how that choice quality relates to socio-demographic characteristics.

We assess the quality of consumers' deductible choices with a simple model that incorporates risk aversion and health risk. We use tools from machine learning to predict health risk as a function of prior utilization and obtain precise predictions for consumers' health spending distributions. We use these risk projections together with our choice model to show that, for values or risk aversion spanning the health insurance choice literature, we can classify individuals into those who should clearly be allocated to a high deductible and those who should clearly be allocated to a low deductible. We quantify tight bounds on the value left on the table due to misallocation, where the bounds are derived from the underlying range of risk aversion parameters that we consider.

Overall, we find that 9% of consumers elect an incremental deductible and that, of those electing an extra deductible, 71% choose the maximum incremental amount of 500 EUR. Our choice model shows that (i) approximately 60% of consumers would be better off if allocated to a higher deductible than the baseline and (ii) almost all of those who would be better off with a higher deductible would be best off with the maximum incremental deductible of 500 EUR. The substantial wedge between what consumers choose and what they would be best off with could be driven by information frictions (see, e.g., Handel and Kolstad (2015*b*) and Bhargava, Loewenstein and Sydnor (2017)) or liquidity constraints (see, e.g., Finkelstein, Hendren and Luttmer (forthcoming)). However, given that the baseline deductible is the default option and consumers have to actively elect an incremental deductible, it is most likely that this large gap is due to default effects (see, e.g., Madrian and Shea (2001) or Handel (2013)). Our forthcoming analysis groups these factors together, studies heterogeneity in this basket of factors, and performs welfare analysis of counterfactual allocative policies that don't require specific knowledge of the underlying determinants of consumer choice errors.

Our primary regression specification studies the impact of a range of socio-demographic factors on deductible choice. The main consumer characteristic that should impact deductible choice in our setting is predicted health risk. We find a positive relationship between good predicted health and incremental deductible takeup conditional on a range of other socio-demographic factors. A consumer that is 100% more likely to have predicted health spending below 375 EUR (i.e. comparing a very healthy to very sick consumer) is approximately 12 percentage points more likely to elect an incremental deductible. We also investigate the impact of health shocks in a panel setup with individual fixed effects and find that health shocks meaningfully impact deductible choice: consumers with a positive health shock are two percentage points more likely to elect a higher deductible after the shock, holding all else equal, while those with a negative health shock are five percentage points less likely to do so.

We next investigate socio-demographic factors that, controlling for health risk, should have minimal to no impact on optimal deductible allocations. Descriptively, we find substantial heterogeneity in deductible choice as a function of educational background. The top five fields, in terms of proportion choosing the 500 EUR

deductible, are (i) statistics (29%), mathematics (21%), physics (21%), architecture (18%), and the physical sciences (18%). The bottom five are literacy and numeracy (2%), occupational health and safety (4%), hair and beauty services (4%), computer use (4%), and child care and youth services (4%). Moving to level of education as opposed to field of education, consumers with advanced degrees above Master's are most likely to elect the extra deductible (16%), while those with a Master's degree (15%) or Bachelor's degree (11%) as their highest degree are next highest. Those with only a high school degree are less likely to elect the higher deductible. We investigate the impact of education field and education level in a regression framework controlling for a rich set of observables (including predicted health expenditures and income). Our estimates show that those with advanced degrees (Masters or above) are approximately 7% more likely than those with high school education or lower to elect a higher deductible. Moreover, when we interact education level with health risk, we find that those with higher education level are making better choices: healthy consumers with high education levels are much more likely to elect the higher deductible (16%) than those with high school education or lower, but sick consumers with high education levels are actually slightly less likely to elect a higher deductible than sick consumers with lower education levels. When we run similar regressions and focus on type of education, our estimates show similar patterns to the descriptive statistics described above. Thus, overall, those with advanced degrees and those in quantitative/analytical fields are most likely to make the correct choice and elect the highest deductible when predicted expenditures are lower. We also investigate the impact of being in a specific profession (rather than having a specific kind of degree) on deductible choice and find that, all else equal, those in more analytic profession (e.g. business services or insurance) are more responsive to predicted health when making deductible choices.

Next, we investigate deductible choice as a function of income and asset holdings. The reason is twofold. First, we study whether income effects or liquidity constraints are a significant determinant of deductible choice. Second, the incidence of choice frictions and health risk may differ across the income distribution, affecting the redistributive consequences of different policy interventions.

Descriptively, as household income increases, the proportion of people who elect an incremental deductible increases, from 5% at low income levels to 14% at high income levels. Since income and health are correlated, we perform this calculation conditional on health risk and a range of other socio-demographic factors. Our estimates show that individual who have higher income are 3 percentage points more likely to choose a high deductible when healthy than those with lower income while they are similarly likely to keep the baseline deductible when sick. We also investigate a specification with individual fixed effects that investigates how individual deductible choices respond to income shocks. We find a precise zero for within-individual income shocks, indicating that income effects are unlikely to be the main driver of the relationship between income and deductible choice.

In addition to income we investigate the impact of wealth, liquid savings, and debt on deductible choices.

Those with higher liquid savings are 2 percentage points more likely to elect an incremental deductible if predicted to be healthy, all else equal. Those with high net worth (top quartile) are 3 percentage points more likely to elect a higher deductible when also predicted to be healthy. Mortgage debt and overall debt have small, statistically significant, impacts on choice, all else equal.

Finally, we analyze the role of peer and location effects on a number of dimensions. Controlling for the range of factors discussed above, we find that if an individual's father (mother) chooses a high deductible, then if that individual has good predicted health they are 19.1% (25.5%) more likely to elect a higher deductible themselves, relative to someone else who is also healthy but whose parents did not elect a higher deductible.

The firm an individual works for also matters for their choice, conditional on the wide range of factors described above. Cross-sectionally, if X% of the other workers in an individual's firm elect a higher deductible, then that individual is .288*X% more likely to elect a higher deductible themselves. We also study a panel AKM-style setup Abowd, Kramarz and Margolis (1999) that studies how movers (e.g. from a firm with many high deductible consumers to one with few) change their purchase behaviors. Preliminary findings show that, all else equal, someone who moves to a firm that has 10% higher deductible takeup is 2.18% more likely to elect a higher deductible themselves (similar in magnitude to the cross-sectional estimate just described). We also investigate location (county) effects and find a positive cross-sectional association between a location's share of choosing a high deductible and a given individual's choice, but preliminary estimates find a precise 0 for this relationship using an AKM-style specification.

Our results paint a detailed picture of the role of socio-demographic factors and peer effects in deductible choice. Taken together, the results show that a range of factors that should have minimal to no impact on what the best deductible allocation in fact have quite strong impacts. We assess the implications of consumers' choices on consumer surplus. Our consumer surplus calculation reflects consumers' certainty equivalents from their deductible choice, conditional on an assumed range of risk aversion coefficients drawn from the literature. It does not reflect the potential consumer surplus impacts of additional cost-sharing on health care utilization (see, e.g., Brot-Goldberg et al. (2017)), though we discuss this aspect of consumer surplus in depth.

First, we consider the consumer surplus gains from an optimal allocation of consumers to deductibles. The average benefit from this policy is in the range [69 EUR, 71 EUR], where the lowest value in the range reflects an assumed CARA coefficient of 10^{-4} and the highest value reflects risk neutrality. We next investigate two mandate policies. The first moves all consumers to the highest policies deductible we observe in sample (875 EUR) and the second moves all consumers to the in sample baseline deductible (375 EUR). Relative to the status quo, mandating that everyone have the 875 EUR deductible decreases consumer welfare by [19 EUR, 21 EUR] while mandating the baseline deductible decreases consumer welfare by 8 EUR. Thus, from a consumer surplus standpoint, a universal mandate is too blunt as an instrument to overcome the information frictions.

We next investigate the equity implications of these policies. We weight outcomes as a function of income using parameters borrowed from the inequality literature (see, e.g., Atkinson (1970)). When we control for health status in the consumer surplus calculation, using income-dependent welfare weight makes the optimal allocation and high-deductible mandate policies slightly more attractive, the former because, controlling for health, lower income is associated with worse in-sample allocations. However, when we don't control for health status (i.e. allow for a correlation between income and health) using income-dependent weights reduces the appeal of the optimal allocation policy, because income is negatively correlated with health so higher income consumers have more to gain from this policy (which shifts people towards higher deductibles generally) on average than poorer consumers. Relatedly, the impact of a high-deductible mandate becomes much worse, and eventually negative, when attaching higher welfare weights to low income consumers. Thus, overall, though lower income consumers make worse choices conditional on health risk, because of the negative correlation with income and health risk they are hurt more by a high-deductible mandate policy than higher income consumers are, even given assumed risk-neutrality. Choice is better than the mandate for low income consumers, even given choice mistakes, but these consumers would be much better off if their choices were better directed (e.g. through smart defaults a la Handel and Kolstad (2015*a*) or Abaluck and Adams (2019)).

Finally, we discuss our results in the context of the moral hazard literature (e.g. Brot-Goldberg et al. (2017), Baicker, Mullainathan and Schwartzstein (2015), Einav, Finkelstein and Schrimpf (2015)) and discuss the consumer surplus, social welfare, and total medical spending implications, given results from the prior literature. We show XXXXX, XXXXX, and XXXXX.[2] We also investigate what the implications for optimal consumer allocations to deductibles would be if the deductible market were a competitive market when adverse selection is a potential concern.[3] We find XXXXXX and XXXXXX.

This paper relates to several distinct literatures. It relates closely to prior work on insurance choice including papers without choice frictions (e.g., Cohen and Einav (2007), Cardon and Hendel (2001)) and with choice frictions (e.g., Handel and Kolstad (2015*b*), Abaluck and Gruber (2016*a*), Abaluck and Gruber (2016*b*), Fang, Keane and Silverman (2008), Ketcham et al. (2012), Bhargava, Loewenstein and Sydnor (2017), Barseghyan et al. (2013)).[4] Relative to the prior literature on insurance choice with frictions, the primary contribution of this paper is to link choice quality to a range of underlying social factors and discuss policymaking in the context of redistribution across different levels of these social factors.

Our analysis also relates to papers that study choice quality in other domains. Most notably, there have been quite a few papers studying choice quality and default effects in retirement savings. For example, Chetty et al.

[2] To be added.

[3] This analysis builds on prior work by Handel (2013), Spinnewijn (2017), Handel, Kolstad and Spinnewijn (2015) and Polyakova (2016).

[4] The literature on insurance choice with choice frictions / behavioral choice foundations is summarized in the chapter by Handel and Schwartzstein (2019).

(2014) study retirement savings in Denmark and show that default effects are much more powerful than subsidies in how they impact consumers' savings portfolios. One key difference between our study and theirs is that we study a context where we can, to a large extent, determine whether consumers are allocated to a poor option for them, while Chetty et al. (2014) show that defaults and subsidies induce very different behavior changes but ultimately it is difficult to tell what savings levels consumers should have in their environment. Cronqvist and Thaler (2004) discuss the privatization of social security in Sweden with a focus on how subtle design factors, such as default effects, can have important implications for consumer choices. Beshears et al. (2016) show that, in a retirement setting, consumers are more likely to switch away from the default option if they will benefit more from doing soon, which is similar to what we find in our environment.

The rest of the paper proceeds as follows. Section 2 describes health insurance in the Netherlands and describes our data. Section 3 presents a simple choice framework and section 4 describes our consumer cost risk prediction model. Section 5 presents our empirical analysis of deductible choice and section 6 presents our analysis of counterfactual policies and consumer welfare. Section 7 concludes.

II Institutional Context and Data

We exploit a unique consumer choice setting in the health insurance market in the Netherlands and link data on health insurance choices to data from various administrative registers. We present here the institutional context and data.

II.A Health Insurance in the Netherlands

All individuals in the Netherlands are by law obligated to buy health insurance from a private insurer.[5][6] Dutch health insurance plans consist of two components. The first component, which is mandatory to buy, is a standardized basic package of health insurance. The basic package fully covers drugs, doctor and hospital expenditures that exceed a deductible. The level of deductible is subject to choice. Individuals have also flexibility to buy this package from any insurer and can choose their level of provider choice (coverage in all hospitals versus a selection). The second component of a health insurance contract, which is not mandatory to buy, is supplementary insurance that can cover dental care, additional physical therapy, alternative medicine, and other care. In 2015, approximately 90% of insurees bought supplementary insurance.

The health insurance market in the Netherlands follows a model of managed competition. The government

[5] An overview of the health system and changes to the health insurance model in the Netherlands can be found in Kroneman et al. (2016). Some of the figures in this section are from Vektis (2015) or homefinance.nl, if we cannot compute them in our data.

[6] This includes children aged below 18 years. These children can freely choose plans among health insurers, do not have to pay for their health insurance plan, and are automatically provided with free dental coverage care. Children also do not face a deductible. Most frequently, children aged below 18 years are insured with the same insurer as (one of) their parents.

strictly regulates the mandatory purchase as well as the contents of the basic package of health insurance, yet allows insurers to compete for consumers on premiums, provider choice, and supplementary insurance. The main rationale for the installment of this model of managed competition in 2006 with the Health Insurance Act of 2006, was to increase competition between health insurers and health providers. In 2015, there were 25 health insurers that together offered 71 separate insurance contracts. Prices for the mandatory health insurance with smallest possible deductible varied from 990 EUR to 1367 EUR with a mean yearly premium of 1133 EUR. The average premium for the supplementary insurance averaged 313 EUR. Consumer choice of a health care plan takes place yearly between a set period between mid November and the end of December for the following year. During that period, health insurers advertise their insurance packages through various media. Relative few people switch insurers (only 6.8% of individuals in 2015).[7]

The consumer choice we focus on in this paper is the choice of deductible. Each individual faces a compulsory deductible of 375 EUR, but can opt for an extra voluntary deductible of 100, 200, 300, 400 or 500 EUR on top of this compulsory deductible, so that the maximum size of the deductible in 2015 is 875 EUR.[8] The compulsory deductible was introduced in 2008, and has gradually increased from 150 EUR in 2008 to 375 EUR in 2015[9], while the options for the extra voluntary deductible have remained the same. By opting for a higher deductible, consumers receive a premium reduction. These premium reductions range by insurance contract, varying from 150 to 300 EUR, with an average premium reduction of 235 EUR. Important to note is that the deductible applies only for certain type of health costs, as some type of expenditures are always fully covered, i.e., not subject to the deductible. These fully insured costs mainly include visits to the general practitioner and maternal care.

The design of the compulsory combined with a voluntary deductible has been a central topic of the policy debate. The desirability of consumer choice in the deductible has last been discussed in the Dutch parliament in 2018. The Minister of Health Affairs argued that having the option of a voluntary deductible increases general support for the health care system in the Netherlands and makes individuals more aware of their health costs.

II.B Data and Sample

We use data on health insurance choices and health expenditures for all individuals in the Netherlands. The data is linked at the Central Bureau of Statistics (CBS) to other administrative registers, which provide information on their income, wealth, education, employment and demographic variables.

[7] This low number might be in part driven by inattention, as by not canceling an insurance contract during this period, a consumer is automatically extending it for the next year. In between the total period of 2006 and 2015, 31% of individuals switched between insurers at least once.

[8] The government does not mandate insurers to provide the choice of voluntary deductible. However, in practice, almost all insurers provide the option take the voluntary deductible; only in 2011 "Zilveren Kruis" does not offer individuals to take an extra deductible for one of its insurance contracts. In 2015, of the 68 insurance contracts that have price information available on https://www.homefinance.nl/zorgverzekering/informatie/2015/zorgpremie-2015.asp, only one insurer (the new entrant ANNO12) does not provide the option to increase the deductible.

[9] The size of the compulsory deductible was 170, 220, 350, 360 and 375 EUR in 2011, 2012, 2013, 2014 and 2015, respectively.

We restrict attention to all individuals who are at least 18 years old in the year they decide on their health insurance contract. In each year, we exclude from the sample adults who have incomplete health data records in the two previous years.[10] The sample consists of about 12 million adults in each year. As explained in Section IV, we use a random sample of 1 million of these individuals to estimate and calibrate a cost prediction model, leaving us with about 11 million adults each year for the analyses.

Health Insurance Deductible Data on health insurance contract choices in the years between 2011 and 2015 are obtained from Vektis, an organization that is responsible for the collection of data from all health insurers. Our data does not include information on the insurer, insurance premium or supplementary insurance. Table 1 shows the take-up of different deductible amounts across the study years. The voluntary deductible take-up in our sample is 8.83% in 2015.[11] More than 2 out 3 individuals opting for an extra deductible take the maximum extra deductible of 500 EUR.

Health Care Costs We also obtain data on health care costs from Vektis. The data contains annual health care expenditures by category. The categories are medicines, hospital care, geriatric care, paramedical care and physiotherapy, mental health care, aids and tools for health, health care in foreign countries, health care transport and other care. In addition to these categories which are subject to the deductible, we also have data on the always-insured types of care. This includes neonatal and maternal care, care by GPs and local nurses, and some other minor items.[12]

Figure 1 presents the distribution of the (log) aggregate health care expenditures that are subject to cost-sharing in 2015. This aggregate distribution is skewed with about 19 percent of individuals make zero expenditures, while for more than 10 percent of individuals expenditures exceed 5000 EUR. Table 3 presents the distributions of annual expenditures for the different categories of medical spending. As expected, these distributions are skewed. Hospital expenditures (1,352 EUR), drugs expenditures (301 EUR) and mental health care (235 EUR) are the three categories with the highest mean spending.

Demographics We obtain information on other variables from various administrative registers. Our data includes standard demographics like age, gender and household status. We use third-party reported information from tax registers on household income and household wealth. The former includes pre-tax income from labor,

[10] As we use two years of data in our prediction model, we exclude individuals with incomplete health data from our sample. Vektis, the provider of the health cost data, also marks the health costs for some individuals as unreliable. This might be due to the fact that these individuals neglected to submit the claims to their insurer, for example when the costs are still under the deductible.

[11] This figure differs considerably from the figure that Vektis advertises in Vektis (2015), but the share has been increasing since its introduction in 2008. This difference is caused by the fact that we only report the deductible take-up for individuals for which Vektis considers the health costs to be reliable.

[12] These other minor items are not measured individually, so they cannot distinguish them from the other insurable costs. Therefore, the measured insurable costs will be slightly larger than the real insurable cost. However, this is likely to be negligible.

self-employment and capital and government transfers. The latter includes information on net worth, liquid and other financial assets, mortgage and other debt. We also observe data on the highest formal education level attained and the specific field of study for individuals who proceed past high school. We use employer-employee data to link individuals' choices at the firm level and sector level. Finally, we can also link individuals' choices to their parents' choices and choices in the neighbourhood at the zipcode level. Table 2 summarises the distribution of the main variables in our sample for the year 2015.

III Choice Framework

In this section we set up a model describing deductible choice, which depends on expected health costs, preferences and potential choice frictions. We study a choice environment with the following features:

- Compulsory deductible of 375
- Choose voluntary deductible d and corresponding premium p from menu $\Omega = \{(d, p_d)\}$
- Individual draws total health cost $x \sim F(x)$

We define expected utility for a fully informed and rational individual as:

$$U_{i,d} = \int u_i(W_i - p_d - s)G_d(s)ds.$$

Here, $G_d(s)$ is the distribution of out-of-pocket spending, conditional on having chosen deductible d, and is derived from $F(x)$ and the relationship $s = min[d, x]$. Using this definition of expected utility, we also define an individual's certainty equivalent from choosing one contract as $CE_{i,d}$, determined as the amount that makes the following equation hold:

$$U_{i,d} = u_i(W_i - CE_{i,d}).$$

Next, we introduce choice frictions to reflect the wedge between actual revealed preference and the preferences of a rational, fully informed individual. We denote individual i's observed choice by C_i and contrast this with the plan choice C_i^* that maximizes her expected utility. The corresponding welfare loss due to choice frictions can be defined as

$$\Delta w_i^* = CE_i^* - CE_i.$$

We will study a set of individual and environmental characteristics such as the level and field of education and deductible choices by parents, colleagues and neighbors, that could be associated with choice frictions in choosing

deductible but likely don't reflect the underlying value one obtains from that deductible.

Empirically, most individuals who elect a deductible higher than the compulsory deductible choose the maximum possible deductible of 875 EUR. In addition, though there is some small variance in the money an individual receives when electing this higher deductible, it is typically right around 250 EUR. Given this, we simplify the contract space to include the following two contracts:

$$\Omega = \{(0,0), (500, -250)\}$$

The choice to take an extra deductible will crucially depend on the probability that health costs do not exceed the compulsory deductible of 375 EUR, which we denote by π_i. Simplifying this further, we can approximate the expected utility by

$$U_{i,d} \approx \pi_i u_i(W_i - p_d) + (1 - \pi_i)\, u_i(W_i - p_d - d).$$

Hence, in expected payoff terms, $\bar{\pi} = 0.5$ is the (approximate) threshold between optimally choosing the additional 500 EUR deductible and saving 250 EUR in premium. We note already that introducing aversion to risk does not increase this threshold by much. For a standard value of absolute risk aversion of 10^{-5} (e.g., cite XX), this threshold increases to 0.5006. Even for extreme risk aversion of 10^{-3}, this threshold is still only 0.56[13].

We will now turn to the prediction of individuals' health expenditures, before studying how their deductible choice depends on the predicted health risk vs. other observable characteristics. We will then demonstrate the robustness of our conclusions to the simplifying assumptions that allow us to focus the analysis.

IV Cost Prediction Model

In order to accurately assess the value of a deductible for every individual, we need to precisely understand each individual's health cost risk. We therefore generate for every individual a prediction of the risk of making low health expenditures with the explicit goal of evaluating the choice of the voluntary deductible. We set up our prediction as a binary classification algorithm that predicts whether consumers will spend more or less than 375 EUR total, which is the minimum deductible they are subject to. If they are predicted to spend less than 375 EUR, then they will be better off with an extra deductible. The rationale for choosing for the binary classification is (1) the simplicity of having one predictive variable that we can use and easily interpret in this paper and (2) the fact that the choice is primarily between taking up an extra deductible or not. From Figure 1, we see that the share of expenditures that fall between 375 and 875 is less than 15%. Furthermore, when predicting an

[13] For a CARA utility function of the form $u(x) = e^{-x\sigma}$, the cutoff value π^* for switching to the high deductible being optimal is given by $(1 - e^{250\sigma})/(e^{-250\sigma} - e^{250\sigma})$.

individual's risk to make expenditures that fall into this region, the sum of the uncalibrated probabilities to end up in any of the intermediary brackets reduces further to below 1%. This provides further support to our focus on the 500 EUR deductible vs. no deductible. We provide further detail on this in Appendix A.C.

For each year, we predict whether or not health spending will be larger than 375 EUR. To obtain such a prediction, we use an ensemble learner consisting of a combination of random forest model, a boosted regression trees model and LASSO. Such an ensemble learner is a standard and often-used machine learning technique for a classification problem. To predicts individual's health costs in the following year, we only include predictors that are known at the time of deciding the deductible (in November and December of the year before the health insurance contract). The predictors that we include are: gender, year of birth, pre-tax household income in deciles $(t-2, t-1)$, working status, education level, education field, and past health spending per category $(t-2, t-1)$. In each year, there are approximately 20 variables for health spending per category, so we have a fine level of detail with which to predict future medical spending. On average, we have approximately 50 predictors in our model.[14]

The actual prediction algorithm follows a couple of steps, similar to the prediction analysis in Einav et al. (2018). First, several key parameters of the prediction models are tuned. Second, the three separate prediction models are trained using the training sample. Third, the obtained predictions are combined into an ensemble predictor. Finally, the ensemble prediction is calibrated. We train the ensemble learner algorithm on a random subsample of 1,000,000 individuals. Using the trained prediction algorithm, we generate predictions for all hold-out observations. All the results and plots in the analyses in this paper are using only this hold-out sample. In Appendix A, we provide more information on the detail of each step of the prediction, as well as a discussion of the results of the prediction model.

Using the ensemble learner, we obtain relatively detailed predictions. The Mean Squared Error (MSE) of our predictions in the hold-out sample amounts to 0.1535, which compares favorably to predictions of a simple OLS model, which results in a significantly worse prediction with an MSE of 0.1860. Figure 2 presents the histogram of the predictions for the ex ante probability of being in the low spending group. There is substantial dispersion in predicted risks over the full range of potential probabilities. The distribution is bimodal, with a substantial share of individuals having either a very low probability or a very high probability of being low spenders. This indicates that health expenses are to a large extent predictable and for many individuals in the population, it should be fairly straightforward to assess whether they are better off choosing a high or low deductible.

[14] We have a different number of predictors in some years, as the categorization of health costs changes slightly in our study period. Every year, we include all health cost categories in our dataset as predictors.

V Deductible Choice

In this section we study the determinants of deductible choice. We relate choice to health risk, which should clearly shift the value a consumer gets from a given deductible choice, as well as other factors that are unlikely to shift the value a consumer gets from that choice. We will look at the determinants of deductible choice with graphical evidence as well as cross-sectional OLS and probit regressions. The OLS regressions have a straightforward specification.[15] [16]

$$Y = \alpha + \beta P(costs < 375) + \gamma X + \epsilon$$

Here, Y is a binary variable that is 1 when an individual takes the 500 voluntary deductible, $P(costs < 375)$ is the predicted probability of having costs lower than 375 EUR, and X includes all variables of interest.

V.A Health Risk

We begin by establishing the relationship between predicted health and deductible choice. Predicted health is the one factor we observe that should clearly impact the value a consumer gets from choosing a certain deductible level. In the following analysis, we always condition on predicted health when studying the impact of other potential choice determinants.

Figure 3 shows the relationship between predicted health risk and deductible choice. Two key facts emerge. First, as expected, people who are healthier are more likely to elect the higher incremental 500 EUR deductible. Second, the relationship between risk and deductible choice is weaker than one would expect if consumers were making value-maximizing choices. For example, a consumer in the healthiest predicted health bin is on average only 14% more likely to elect the high deductible, though the simple choice exercise conducted in Section III suggests that this difference should be much larger if consumers are value-maximizing.

The same two key facts are confirmed when using only within-individual variation in predicted health risk. This is shown in Table 11. Increases in health risk lead to statistically significant decreases in extra deductible take up, but the responses are limited in size. The first-difference estimates also indicate that individuals are more responsive to negative (relative to positive) changes and somewhat more responsive to large (relative to small) changes.

[15] In the specifications with interactions, the specification is: $Y = \alpha + \beta P(costs < 375) + \gamma X + \nu P(costs < 375)X + \epsilon$, where the extra term $P(costs < 375)X$ is the cost probability interacted with the other variables of interest.

[16] Using a probit model, the specification is: $Pr(Y = 1) = \Phi\left(\alpha + \beta P(costs < 375) + \gamma X + \epsilon\right)$.

V.B Income, Financial Status and Demographics

Figure 4 plots the fraction who take up the incremental deductible as a function of their gross income (including capital income and government transfers). At first sight, income effects are as important in magnitude as health risks, but once we control for predicted health and other variables capturing socio-economic status, the gradient is much smaller. This is also confirmed in Table 4 showing how income relates to deductible choice as a function of predicted health status, controlling for age, gender, education level, and household structure. Relative to those in the lowest income quartile, those in the top quartile are 2.1% more likely to elect the higher deductible (column 1). However, the high-income quartile seems more responsive to their underlying health risks. Panel A in Figure 5 shows how take up of the incremental deductible changes with the underlying health risks for the different income quartiles. It is clear that those with higher income are more likely to take up the incremental deductible as they are predicted to be healthier.

We further assess the potential role of income effects. We first investigate how deductible choice responds to income shocks. Table 12 indicates that deductible choices do not significantly vary with within-individual income variation, both when estimated using individual fixed effects or using first-differences. This is suggestive that the correlation between income and predicted health is not driven by income effects.

In addition to income, we observe a range of variables related to an household's financial status. We study:

- Wealth: the household's net worth, which we split each household into population quartiles.
- Debt: we look at whether the household has mortgage debt or any other debt (study and other debt).
- Liquidity: a measure of liquid savings, equal to 1 if a household has more than 2000 EUR in liquid savings and 0 otherwise.

Table 5 presents the results of a regression examining the association between these financial variables and incremental deductible take up, controlling for predicted health spending, age, income, gender, household structure, and education level. We find that household liquid savings are positively correlated with deductible take up: having liquid savings of greater than 2000 EUR is associated with a 1.2 percentage point increase in deductible take up. Note that in theory, liquidity and debt constraints could either increase the demand for insurance (to avoid large expenditures) or reduce the demand for insurance (to avoid paying the premium). The sign of the effect we find is consistent with the former explanation. In line with this, we also find that households who are in debt (excluding mortgage debt) are also less likely to take-up the deductible. The effects, however, are small in both cases. Finally, we find that take up rate for wealthier individuals is higher as well and this effect is fully driven by wealthier individuals with better health. That is, wealthier individuals are more responsive to taking the incremental deductible as they become healthier. Hence, rather than capturing wealth effects on insurance

choices, this result could be indicative of information frictions for people with less financial resources.

Table 4 finally shows that there are also relatively modest average effects of age, gender and household composition on deductible choice, once controlling for health risk, income and education level. While this could reflect some minor differences in preferences, there are again significant differences in responsiveness to underlying health risks. For example, when predicted to be healthy, the young are significantly more likely to elect a higher deductible than the old. Again, this could be indicative of the potential role played by information frictions, which we turn to next.

V.C Education and Expertise

We now study the relationship between deductible take-up and education level and field of expertise. This allows us to probe into the potential role of cognitive skills and information frictions in insurance choices.

Figure 4 plots the relationship in the data between education level and deductible choice. There is a clear relationship where higher levels of education are associated with better deductible choices. Those in the healthiest predicted risk decile with Masters degrees or other advanced degrees elect the higher deductible about 25% of the time, while those with less than high school education in the healthiest predicted decile do so approximately 6% of the time and those with high school education only, in the healthiest decile, do so approximately 12% of the time. For all of these education levels, when people are predicted to be sick they almost never elect the higher deductible.

Table 4 shows the impact of education level on deductible choice controlling for health status, gender, age, income, and family structure. The results are similar in spirit to the descriptive relationship described in figure 4: conditional on having predicted total spending below the threshold where a risk-neutral individual should choose the high-deductible, someone with a Masters degree is 5.8% more likely than someone with only high school education to actually elect the high deductible, someone with another kind of advanced degree is 7.2% more likely, and someone with a bachelor degree is 3.4% more likely. This difference is essentially all due to the difference in slope of deductible choice as a function of health status: these groups have similar predicted deductible choice probabilities when sick (close to 0%), but quite different when healthy.

Figure 4 plots the relationship in the data between education field and deductible choice. Since there are many education fields, the figure plots the relationship between deductible choice and predicted health risk for 6 specific fields. Statistics majors are the most responsive to predicted health risk: they choose the additional deductible almost 50% of the time when they are in the healthiest predicted deciles and choose the additional deductible almost never when they are in the sickest predicted deciles. The figure illustrates that, for professions where people are likely to be more analytical in nature and professions that require more advanced schooling,

deductible choice looks more like we would expect for a fully-informed rational consumer. The figure also shows, e.g., that people who study hair and beauty services have limited responsiveness of deductible take up to predicted health status.

Table 6 illustrates this link more broadly, presenting key statistics related to education field and deductible choice. More quantitative fields, grouped at the top of the table, exhibit greater responsiveness to predicted health risk when making deductible choices while those in less quantitative fields, grouped at the bottom of the table, exhibit lower responsiveness.

Table 7 presents regression analysis for choice of deductible by education field, controlling for predicted health risk, income, age, education level, and household structure. For parsimony, we include the six professions included in Figure 4 in the regression analysis. The results paint a compelling picture: even controlling for these other factors, more quantitative / analytic professions (e.g. statistics) are more responsive to predicted health when making deductible choices. For example, someone with statistics training who has low predicted health risk is 13.7% more like to choose a higher deductible, controlling for age, income, gender, and education level than someone with hair and beauty training.

In addition to studying field of education, we study professional sector. While this, of course, is correlated with field of study, there is some differentiation between field or study and what someone chooses to do professionally.

Figure 4 illustrates our relationship of interest for six sectors: business services, retail, insurance and health insurance, construction, cleaning, and public utilities. A similar pattern emerges to that found when classifying by educational field: the professional sectors that involve more day-to-day analytical work and require more educational training are also more likely to have a stronger relationship between predicted health risk and deductible choice. Of the six professions shown in the figure, business services and insurance exhibit the strongest responsiveness (0% w/ high deductible when predicted to be sick, around 20% when predicted to be healthy). Cleaning exhibits the lowest responsiveness (0% w/ high deductible when predicted to be sick, around 7% when predicted to be healthy).

Table 8 illustrates this link more broadly, presenting key statistics related to professional sector and deductible choice. More analytical sectors, grouped at the top of the table, exhibit greater responsiveness to predicted health risk when making deductible choices while those in less analytical sector, grouped at the bottom of the table, exhibit lower responsiveness.

Table 9 presents regression analysis for choice of deductible by professional sector, controlling for predicted health risk, income, age, education level, and household structure. For parsimony, we include the six sectors included in Figure 4 in the regression analysis. The results show that, controlling for these other factors, more analytic professions (e.g. business services and insurance) are more responsive to predicted health when making deductible choices. For example, someone in the insurance sector who has low predicted health risk is 5% more

like to choose a higher deductible, controlling for age, income, gender, and education level, than someone in the public utilities sector.

V.D Choices by Parents, Colleagues and Neighbors

The prior analyses are tied to larger classifiers (e.g. income, education level, profession) that tell a lot about an individual's background or lifestyle. In this section we investigate environmental factors, which also vary individual by individual conditional on these other classifiers. Specifically, we investigate how deductible choice depends on (i) the deductible choices of one's parent and (ii) the firm one is employed by and (iii) the deductible choices made by others in the area someone lives.

For parental deductible choice, we use a variable that is one if the parent has elected a 500 EUR extra deductible in the same year. For firm-specific share we use the proportion of individuals who have taken up the extra deductible in the same firm. For location-specific share, we use the proportion of individuals in an individual's 6-digit postcode who have taken up the higher deductible.

Table 10 presents the results of our regression analysis studying the impact of the choices by parents, co-workers and individuals living in the same area for an individual's deductible choice. The analysis controls for health risk, education level, age, gender and income.

We find meaningful effects of parental deductible choices on their child's choice. If an individual's father (mother) chose a higher deductible, and that individual has good predicted health, then that individual is 19.1% (25.5%) more likely to elect the higher deductible themselves, relative to someone who is healthy and whose parents did not elect the higher deductible (and holding a range of other factors constant).

In additional to this cross-sectional analysis, we implement a panel analysis to investigate the impact of lagged parent deductible choices on future child choices. Currently, our preliminary analysis indicates that parental choices have a large effect on children. Overall, when a parent elects a deductible, in the next year their child is (i) more likely to elect a deductible and (ii) even more likely to elect a deductible if the parent did previously and won financially, i.e. spent low enough that, ex post, the higher deductible turned out to be the better choice.

We also investigate the relationship between firm share of individuals electing an extra deductible and individual deductible choice. We do this in both cross-sectional regressions and panel AKM-style regressions. The cross-sectional regression analysis shows that, controlling for a range of factors, when the share of those in a firm choosing an extra deductible is X% higher, the probability a given individual in the firm chooses an extra deductible is .288*X% higher, as seen in Table 10.

In order to give this association a causal interpretation, we run an AKM-style regression leveraging individuals

who move across firms to identify firm-specific effects. We estimate:

$$Y_{i,t} = \alpha_i + \gamma_f + \varsigma_z + X_{i,t} + \epsilon_{i,t}$$

Here, α_i is an individual fixed effect, γ_f is a firm fixed effect and ς_z is a location fixed effect. $X_{i,t}$ reflects individual-specific changes over time in income and health status. This setup identifies firm fixed effects using individuals who switch firms and identifies location fixed effects using movers.

Figure 6 shows the impact of the share in a firm electing a higher deductible on an individual's likelihood of electing a higher deductible. The causal estimate is 0.218, indicating that for every 10 percentage point rise in firm share with an extra deductible, an individual is 2.18 percentage points more likely to elect that deductible, all else equal. The figure also plots the relationship between the estimated firm fixed effect and the observed fraction of individuals in the firm taking up an additional deductible. These fixed effects closely track the observed share of deductible choice by firm, suggesting that, in this instance, there is limited difference in the cross-sectional association relative to the causal panel estimate.

For location-specific effects, the cross-sectional analysis in Table 10 finds that the fraction of people in one's location choosing to elect a higher deductible has a meaningful positive association with one's own choice. If an individual is predicted to be healthy, then a 20 percentage point increase in a location's high deductible share is associated with a 12.7% increase in that individual's likelihood of electing a higher deductible, holding all else equal.

However, in the preliminary AKM-style analysis, we find no causal effect of the share in a given location electing a higher deductible on an individual's likelihood of electing a higher deductible. Panel B[17] of Figure 6 shows that the relevant estimate is small in magnitude and statistically insignificant and that the estimated fixed effects are not meaningfully correlated with the observed share in a location electing a higher deductible.

VI Choice Frictions, Welfare and Policy

We now study the efficiency and equity implications of regulated deductible choice in the Netherlands. We first characterize the potential financial savings from improving individuals' deductible choices and study how the incidence of choice frictions falls on individuals with different observable characteristics. In addition to caring about efficiency, a regulator might care about the equity implications of deductible choice and weight characteristics such as income and health risk in evaluating the social impact of these counterfactual policies. We investigate the value of offering the deductible choice when accounting for equity and also evaluate counterfactual

[17] Figure to be added.

mandates which do not leave any choice to individuals.

VI.A Expected Cost Savings

We calculate the potential cost savings for each individual by comparing her deductible choice C_i to the choice that minimizes her expected out-of-pocket expenditures. These cost savings simply correspond to the difference in certainty equivalents for risk-neutral preferences:

$$\Delta w_i^{*,\sigma=0} = CE_i^{*,\sigma=0} - CE_i^{\sigma=0}.$$

In our stylized model in Section III, it is optimal (in expected payoff terms) for an individual to choose the extra 500 EUR deductible if the predicted probability to make lower costs is higher than 50 percent.[18] On average, we find that 6.4% of individuals take up the extra deductible, while 60.3% of individuals would benefit from it in expectation. Overall, 45.3% of individuals choose the cost-minimizing deductible.[19]

The expected costs savings combine the expected loss from over-insurance for low-cost individuals ($\hat{\pi} \geq .5$) who do not take the extra deductible and from under-insurance for high-cost individuals ($\hat{\pi} < .5$) who do. The bin-scatter plot in Figure 7 shows how the cost savings vary with the predicted probability $\hat{\pi}$. This graph reflects both the mechanical relationship between $\hat{\pi}_i$ and the potential cost savings, but also the actual distribution of choices made conditional on a given predicted $\hat{\pi}_i$, as shown in Figure 3. On the one hand, very few individuals under-insure: most individuals with high predicted risk stick to coverage without extra deductible, as they should. On the other hand, relatively few individuals opt for the deductible when they should and the expected loss from over-insurance increases as the predicted risk is lower.

A welfare analysis based on expected cost savings ignores risk aversion as a potential driver of deductible choice. However, given the present financial stakes, introducing risk aversion does not change any of the conclusions. Even with substantive risk aversion, for example assuming CARA preferences $u(c) = -exp(-\sigma c)/\sigma$ with $\sigma = 10^{-4}$ (e.g., Handel and Kolstad (2015*b*)), the threshold probability to take up the deductible increases only from .5 to .506. Our findings[20] show that the potential welfare gains Δw_i^* for risk-averse and risk-neutral preferences are basically the same. To rationalize the low take-up rate, we would need extreme curvature in preferences ($\sigma = 10^{-2}$). In a similar spirit, we can calculate the switching costs that would rationalize the observed choices, assuming that taking no extra deductible is the default option.[21] Despite the moderate financial stakes, the required switching

[18] The expected loss equals $250 - (1 - \hat{\pi}_i)500$ if an individual does not opt for the extra deductible. This is positive if the predicted probability to make low costs is higher than 0.5.

[19] For the analysis in this section, we remove the small percentage (2.6%) of consumers who elect positive incremental deductibles less than the 500 EUR maximum.

[20] Figure to be added.

[21] We assume a simple additive switching cost to choosing $d = 500$, so that the expected gain from take the extra deductible equals $250 - (1 - \pi) * 500 - s$.

costs amount to 191.4 EUR. The reason is that the take-up rate only picks up slowly as individuals get more healthy. The question also remains whether heterogeneity in extreme risk aversion or large switching costs is a plausible driver of the variation in deductible choices across individuals with different observable characteristics.

VI.B Heterogeneous Choice Frictions

Our earlier findings indicated the key role of income, education and information, but also the exposure to peers in explaining the difference in deductible take-up and how it relates to predicted health risk. To gauge the corresponding welfare effects more systematically, we analyze the expected losses for groups with different characteristics. As discussed, the potential cost savings are closely linked to the predicted health risks - individuals with worse health choose better. However, this relation may be entirely driven by a confounding default effect causing individuals not to take a deductible, which happens to be the correct decision for individuals with worse health. To account for this, we control for differences in health by comparing the quality of the decisions by different groups assuming the population distribution of health risk for each group.

We run a linear model of deductible choice $S_{i,d}$ on predicted health risk, a rich set of observables, and interactions between the two,[22]

$$\begin{aligned} S_{i,d} &= S_d(X_i, \pi_i) + \epsilon_i \\ &= \sum_d \alpha_d \cdot X_i^d + \sum_{\delta_\pi} \alpha_{\delta_\pi} \cdot 1[\pi_i \in \delta_\pi] + \sum_d \sum_{\delta_\pi} \alpha_{d,\delta_\pi} \cdot X_i^d \cdot 1[\pi_i \in \delta_\pi] + \epsilon_i \end{aligned}$$

We use population health deciles and interact each demographic category separately with the health deciles. In practice, this results in approximately 1100 parameter estimates.

We then use the results from this regression to predict the deductible choices for each individual demographic vector X_i under the population health distribution. That is, we define

$$\tilde{S}_d(X_i) = \int_\pi \hat{S}_d(X_i, \pi)\, dF_\pi,$$

which gives us the predicted deductible take-up rate for each observed X_i combination but as if there were a population of individuals with that X_i with the same health distribution as the overall population. In the same way, we predict the choice quality for individuals with demographic vector X_i, as captured by the probability to choose the contract that minimizes expected expenditures, $\tilde{S}^*(X_i)$, and the corresponding average financial loss

[22] Following our regression analysis, we partition the population based on age (4 groups), gender (2 groups), partner (2 groups), children (2 groups), education level (4 groups), education field (102 groups), occupation (6 groups), sector (70 groups), mother's deductible choice (3 groups - incl. not observed), father's deductible choice (3 groups - incl. not observed), avg. deductible choice in firm (continuous, plus an indicator for whether observed), avg. deductible choice in neighborhood (continuous, plus an indicator for whether observed)

$\tilde{w}^*(X_i)$.[23] The choice quality varies through the deductible choice predicted by the set of demographics X_i for different health risks, but again reflects the population distribution of health risks.

Figure 8 compares the distribution of predicted deductible choice, with and without the effect of healthcare cost risk. These are denoted in previous equations as $\hat{S}(X_i, \pi_i)$ and $\tilde{S}(X_i)$ respectively. As shown before, health has a meaningful impact on deductible choice, but there is substantial heterogeneity in likelihood of choosing a deductible just as a function of X_i, netting out health effects. Figure 9 illustrates the corresponding heterogeneity in choice quality, plotting the distribution of the probability to choose the cost-minimizing deductible and of the expected losses from sub-optimal choices. These are denoted by $\tilde{S}^*(X_i)$ and $\tilde{w}^*(X_i)$ respectively. While losses range up to 200 EUR when factoring health risk into choices, when assuming the population distribution of health for a given X_i the expected loss ranges between 50 and 80 as a function of X_i. This variation corresponds to substantial heterogeneity in choice quality. The very best decision makers choose the cost-minimizing deductible 86% of times, but less than 1% have a probability higher than 70% to make the right choice, which seems still quite far from how a *homo economicus* would behave. 93.7% of individuals, however, are predicted to choose worse than someone who randomizes. 25% are predicted to get it wrong at least 57.6% of times. These substantial choice frictions are mostly driven by over-insurance. We see that individuals with observable characteristics who are more likely to take a deductible overall are more responsive to their predict risk as well. The relationship between the take-up rate and the choice quality itself is close to linear as shown in Figure 10.

Table 13 compares the demographics of the individuals bearing the largest and smallest expected loss, corresponding to the observations contained in the rightmost and leftmost points in the left panel of Figure 9. The 5% best decision makers are typically younger, men, with higher income and higher education. In the group of worst decision makers, three out of four are high school drop-outs (only 7% in the group of best decision makers are high school drop-outs). The opposite is true for the highly education individuals, who are basically not represented in the group of worst decision makers. The patterns by field of education and occupation are being confirmed here again.

Figure 11 shows the responsiveness of deductible choices to differing health risk within groups of similar choice quality. We can see again that those who bear smaller losses are not only more likely to take up the deductible overall, but their deductible choice is also much more responsive to their own underlying healthcare cost risk. The performance of the very best decision makers is quite striking relative to the others. While the median basically

[23] To be precise, we have

$$\begin{aligned}
\tilde{S}^*(X_i) &= \int_\pi \{1[\pi \leq .5]\hat{S}_0(X_i, \pi) + 1[\pi > .5]\hat{S}_{500}(X_i, \pi)\} dF_\pi, \\
\tilde{w}^*(X_i) &= \int \{1[\pi \leq .5]\hat{S}_{500}(X_i, \pi)\left[CE^{\sigma=0}_{\pi,0} - CE^{\sigma=0}_{\pi,500}\right] + 1[\pi > .5]\hat{S}_0(X_i, \pi)\left[CE^{\sigma=0}_{\pi,500} - CE^{\sigma=0}_{\pi,0}\right]\} dF_\pi.
\end{aligned}$$

We use the average predicted risk for the different health deciles to calculate the certainty equivalents and to determine whether one should take up the deductible or not.

sticks to the compulsory deductible regardless of the underlying health risk, the take-up rate of the top 1% of decision makers comes close to the 45-degree line. The take-up rate increases from 0 to more than 60 percent as the probability of low health care costs increases from 0 to 75 percent.

VI.C Welfare and Counterfactual Policies

In addition to caring about efficiency, a regulator might care about the equity implications of deductible choice. A key equity concern is the incidence of choice frictions along the income distribution. Figure 12 shows how the choice quality varies with income, assuming the population distribution of health status for each income decile.[24] Individuals with higher income make better choices and thus experience lower losses than those those with lower income. At the top end of the income distribution individuals are 10 percent more likely to choose the optimal deductible than at the bottom of the distribution. The regulation allowing for deductible choice thus seems regressive. This comparison, however, controls for health risk. Panel A of Figure 4 already showed the role of health risk and income together, plotting the take-up rates by health deciles for different income groups. The largest losses are indeed borne by the individuals with the best health and lowest income. So while less healthy individuals are somewhat protected against making bad choices in this choice environment, the low-income individuals are not, especially not when in good health.

We calculate the overall cost savings from eliminating choice frictions, as reported in Column 1 of Table 14. To reflect that a social planner may want to weight characteristics such as income and health risk in evaluating the social impact of one policy relative to another, we report the cost savings when introducing two levels of inequality aversion through income-dependent weights, and when controlling for health risks. In the former scenario, we weight the cost savings for each income decile y_δ by $y_\delta^{-\epsilon} / \left(\bar{y_\delta^{-\epsilon}}\right)$ for $\epsilon = .5$ and $\epsilon = 1.5$.[25] In the latter scenario, we assume again the population distribution of health status for each income decile.[26] The unweighted average cost savings from eliminating choice frictions are 70.7 EUR. While small in absolute terms, these savings are very substantial as individuals have only 143 EUR at stake on average.[27] The welfare loss due to choice frictions is, however, smaller when giving relative more weight to individuals with lower income. These individuals tend to gain less from the optimal choices due to their worse health on average. The opposite is true when controlling for health risks (Column (4)).

To further investigate the impact of the opportunity to elect an additional deductible, we also evaluate

[24] For this analysis, we predict deductible choices depending on the income and health deciles only.

[25] With this, we follow Atkinson (1970), where the Atkinson index of inequality uses a social welfare function of the form $y_i^{1-\epsilon}$, and where $\epsilon \geq 0$ is a measure of inequality aversion. In the present case, each decile's average income y_i is therefore multiplied by $y_i^{-\epsilon}$, and standardized by the average $\bar{y_i^{-\epsilon}}$ to ensure comparability with the unweighted case.

[26] We could also model equity concerns for the social planner more generally using a function H that differentially weights outcomes for individuals as a function of their characteristics X_i, $H(CE_{i,d})$. Note that we don't discuss the welfare impact of a higher or lower deductible that results from consumer moral hazard of consumer behavioral hazard. We will discuss the welfare impact of consumer price sensitivity later.

[27] We define the stake as $|250 - (1 - \hat{\pi})500|$ EUR, which is at most 250 EUR and equal to 0 for individuals with $\hat{\pi} = .5$.

counterfactual policies that require all individuals to take either the 500 EUR deductible (and get a premium reduction of 250 EUR, see Columns (2) and (5)) or take no extra deductible at all (and get no premium reduction, see Columns (3) and (6)). Importantly, both universal mandate policies would decrease welfare. Health care costs are predictable and thus the optimal choice varies substantially across individuals. Even individuals often do not make the optimal choice, a universal mandate into the same coverage for everyone seems too blunt an instrument to overcome the information frictions. The welfare loss when mandating everyone to take the extra deductible is particularly large when equity concerns are strong (−121.50 EUR). The reason is again that low-income individuals have worse health on average and lose more from being forced to take a higher deductible. The opposite is again true once we control for health risk, reflecting the higher incidence of choice frictions among low-income individuals. This incidence effect is small though relative to the effect of the positive correlation between income and health. For completeness, we perform the same computations for a range of risk aversion levels. The conclusions are robust to the introduction of even very high levels of constant absolute risk aversion ($\sigma = 10^{-3}$).

VII Conclusion

To be added.

References

Abaluck, Jason, and Abigail Adams. 2019. "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses." Yale working paper. 5

Abaluck, Jason, and Jonathan Gruber. 2016*a*. "Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance." *American Economic Review*, 106(8): 2145–84. 5

Abaluck, Jason, and Jonathan Gruber. 2016*b*. "Improving the Quality of Choices in Health Insurance Markets." NBER Working Paper No. 22917. 5

Abowd, John, Francis Kramarz, and David Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251–333. 4

Atkinson, Anthony. 1970. "On the Measurement of Inequality." *Journal of Economic Theory*, 2: 244–263. 5

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. 2015. "Behavioral Hazard in Health Insurance." *Quarterly Journal of Economics*, 130(4): 1623–1667. 5

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua Teitelbaum. 2013. "The Nature of Risk Preferences: Evidence from Insurance Choices." *American Economic Review*, 103(6): 2499–2529. 5

Beshears, John, James Choi, David Laibson, Brigitte Madrian, and Sean Wang. 2016. "Who is Easier to Nudge?" Harvard University Working Paper. 6

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor. 2017. "Choose to Lose: Health Plan Choices from a Menu with Dominated Option." *The Quarterly Journal of Economics*, 132(3): 1319–1372. 2, 5

Brot-Goldberg, Zarek C, Amitabh Chandra, Benjamin R Handel, and Jonathan T Kolstad. 2017. "What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." *The Quarterly Journal of Economics*, 132(3): 1261–1318. 4, 5

Cardon, James, and Igal Hendel. 2001. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *The RAND Journal of Economics*, 32(3): 408–427. 5

Chetty, Raj, John N. Friedman, Søren Leth-Petersen, Torben Heien Nielsen, and Tore Olsen. 2014. "Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark." *The Quarterly Journal of Economics*, 129(3): 1141–1219. 1, 5, 6

Cohen, Alma, and Liran Einav. 2007. "Estimating Risk Preferences from Deductible Choice." *The American Economic Review*, 97(3): 745–788. 5

Cronqvist, Henrik, and Richard Thaler. 2004. "Design Choices in Privatized Social Security Systems: Learning from the Swedish Experience." *American Economic Review*, 94(2): 424–428. 6

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2015. "The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D." *Quarterly Journal of Economics*, 130(2): 841–899. 5

Einav, Liran, Amy Finkelstein, Sendhil Mullainathan, and Ziad Obermeyer. 2018. "Predictive Modeling of U.S. Health Spending in Late Life." *Science*, 360(6396): 1462–1465. 11, 53, 54

Enthoven, Alain, Alan Garber, and Sara Singer. 2001. "Near-Universal Coverage Through Health Plan Competition: An Insurance Exchange Approach." *Covering America: Real Remedies for the Uninsured, Editors J. Meyer and E. Wicks*, Washington DC: 155–172. 1

Fang, Hanming, Michael Keane, and Daniel Silverman. 2008. "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market." *Journal of Political Economy*, 116(2): 303–350. 5

Finkelstein, Amy, Nathan Hendren, and Erzo Luttmer. forthcoming. "The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment." *Journal of Political Economy*. 2

Handel, Benjamin. 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." *American Economic Review*, 103(7): 2643–2682. 2, 5

Handel, Benjamin, and Jonathan Kolstad. 2015*a*. "Getting the Most From Marketplaces: Smart Policies on Health Insurance Choice." *Brookings Hamilton Project Discussion Paper 2015-08*. 5

Handel, Benjamin, and Jonathan Kolstad. 2015*b*. "Health Insurance For Humans: Information Frictions, Plan Choice, and Consumer Welfare." *American Economic Review*, 105(8): 2449–2500. 2, 5, 18

Handel, Benjamin, and Joshua Schwartzstein. 2019. "Behavioral Economics and Health Care Markets." *Handbook of Behavioral Economics*. 5

Handel, Benjamin, Jonathan Kolstad, and Johannes Spinnewijn. 2015. "Information Frictions and Adverse Selection: Policy Interventions in Health Insurance Markets." NBER Working Paper No. 21759. 5

Hastings, Justine, Chris Neilson, Anely Ramirez, Unika Shrestha, and Seth Zimmerman. 2013. "(Un)informed College Choice: Evidence From Chile." Brown University, Working Paper. 1

Ito, Koichiro. 2015. "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *American Economic Review*, 104(2): 537–563. 1

Ketcham, Jonathan, Claudio Lucarelli, Eugenio Miravete, and Christopher Roebuck. 2012. "Sinking, Swimming, or Learning to Swim in Medicare Part D?" *The American Economic Review*, 102(6): 2639–2673. 5

Kroneman, Madelon, Wienke Boerma, M van den Berg, P Groenewegen, J de Jong, and E. van Ginneken. 2016. "Netherlands: Health System Review." *Health Systems in Transition*, 18(2): 1–240. 6

Madrian, B., and D. Shea. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *Quarterly Journal of Economics*, 141(4): 1149–1187. 2

Neilsen, Christopher. 2017. "Targeted Vouchers, Competition Among Schools, and the Academic Achievements of Poor Students." Princeton Working Paper. 1

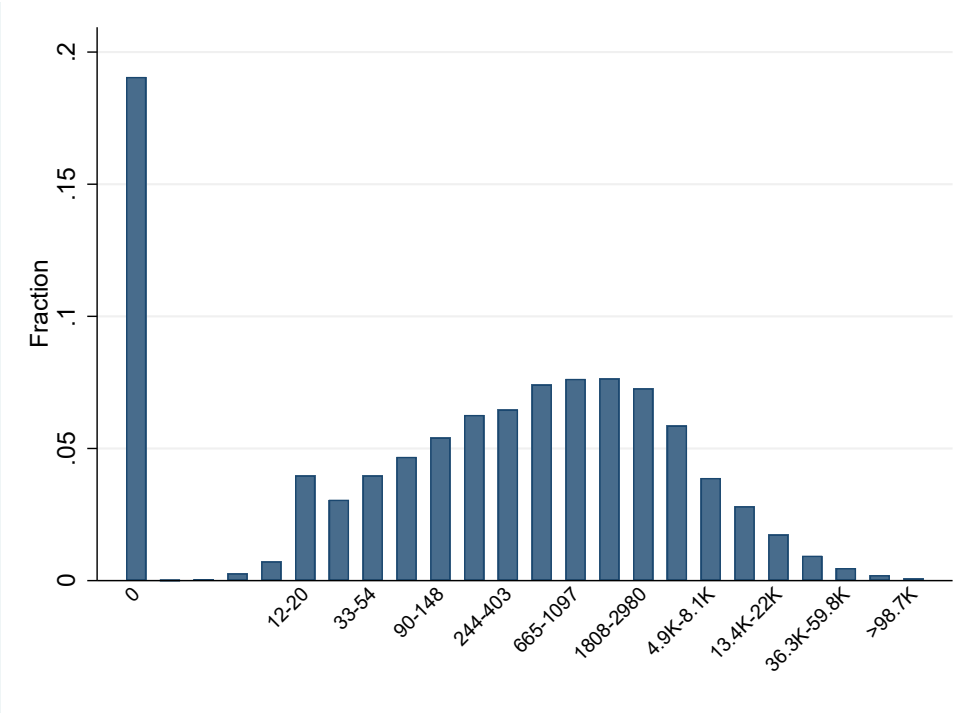
Polyakova, Maria. 2016. "Regulation of Insurance with Adverse Selection and Switching Costs: Evidence from Medicare Part D." *American Economic Journal: Applied Economics*, 8(3): 165–195. 5

Spinnewijn, Johannes. 2017. "Heterogeneity, Demand for Insurance and Adverse Selection." *American Economic Journal: Economic Policy*, 9(1): 308–343. 5

Vektis. 2015. "Zorgthermometer. Verzekerden in Beeld 2015." Vektis Publicaties. 6, 8

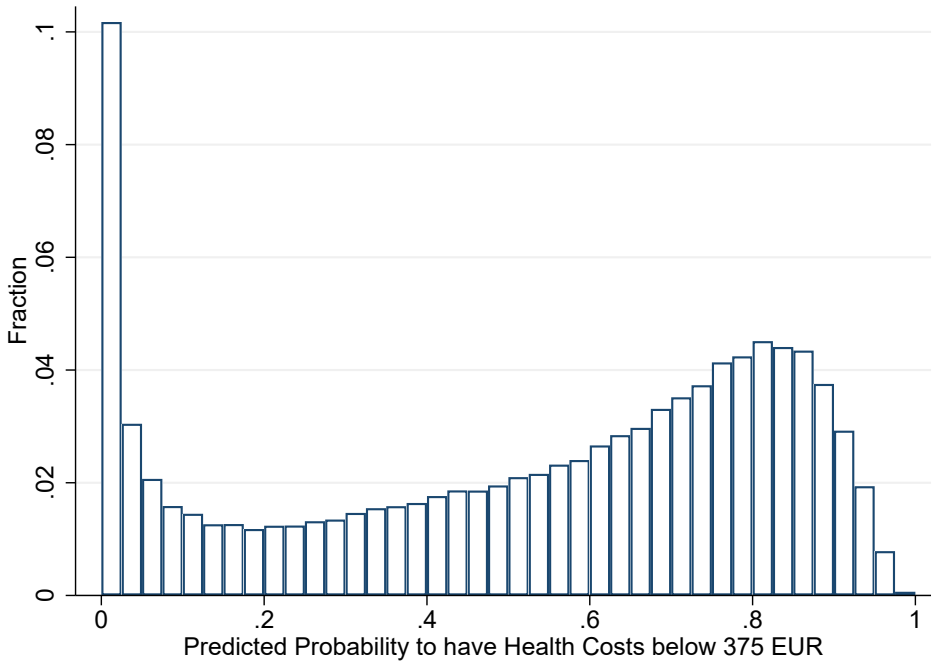
Figures

Figure 1: Distribution of Insurable Health Care Costs



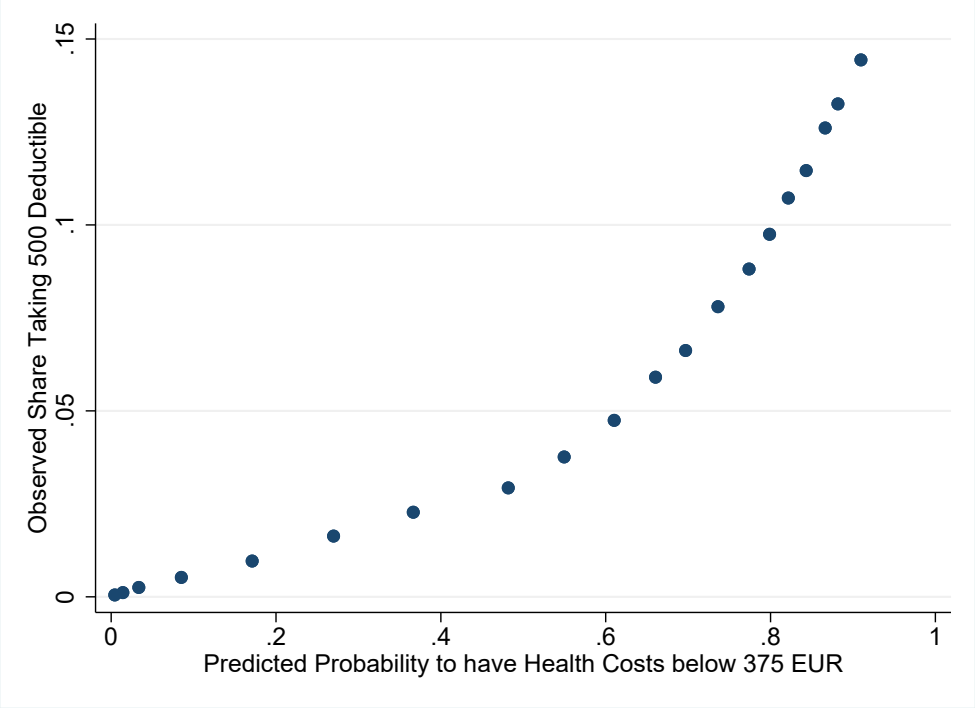
Notes: This figure shows the distribution of the total yearly insurable health care costs in 2015 for all individuals in our baseline sample. The graph uses a log scale: the first bin covers 0 costs, the second bin covers costs from exp(0) to exp(0.5), the third covers costs from exp(0.5) to exp(1), and so on. The labels on the x-axis show the corresponding EUR amounts of selected bins.

Figure 2: Distribution of Cost Probability Predictions



Notes: This figure shows the distribution of the predicted probabilities of having health costs below 375 EUR. These probabilities are obtained when predicting the binary variable (having insurable health costs below 375) with the ensemble machine learner described in Section III, and further in Appendix A. From this figure it can be seen that our model predicts approximately 10% of individuals to have less than 2.5% chance to have health costs below 375 EUR. In contrast, the fraction of individuals for which our model predicts a 97.5% or higher probability to have costs below 375 EUR is very small, which is consistent with the notion that accidents and unexpected costs happen to anyone with some probability.

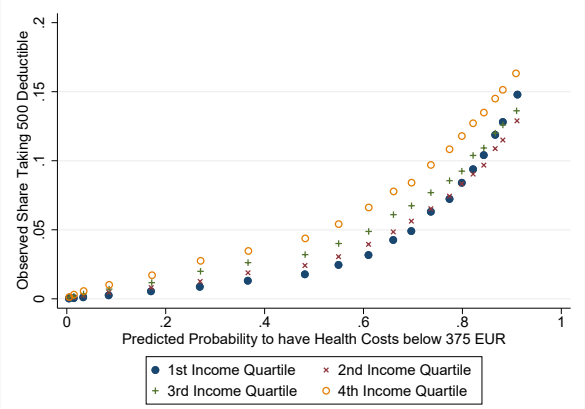
Figure 3: Take-up of 500 Voluntary Deductible as Function of Predicted Health Costs



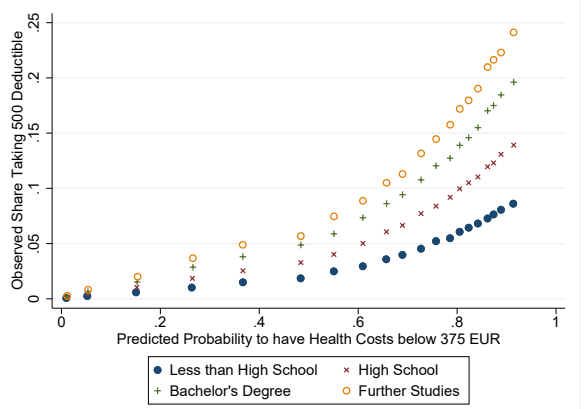
Notes: This figure is a binned scatterplot of the relationship between the take-up of the predicted probability of having costs below 375 EUR (staying under the voluntary deductible range) and the take-up of the voluntary 500 EUR extra deductible.

FIGURE 4: TAKE-UP OF 500 VOLUNTARY DEDUCTIBLE AS FUNCTION OF PREDICTED HEALTH COSTS

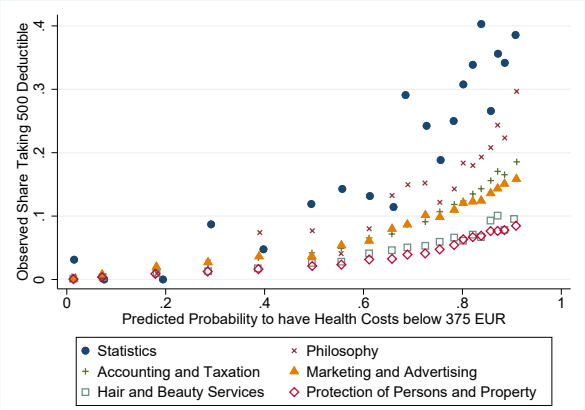
A. By Gross Income Group



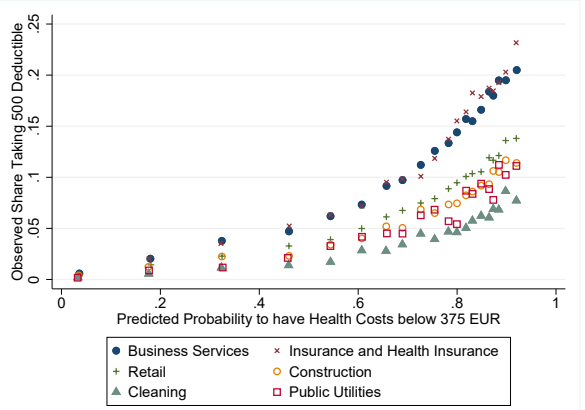
B. By Level of Education



C. By Field of Study

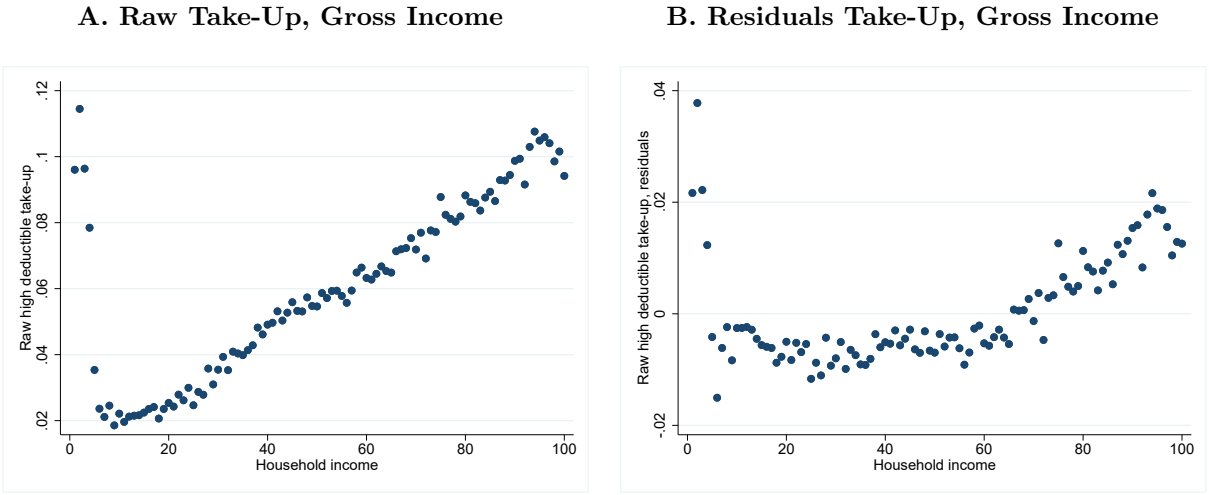


D. By Sector



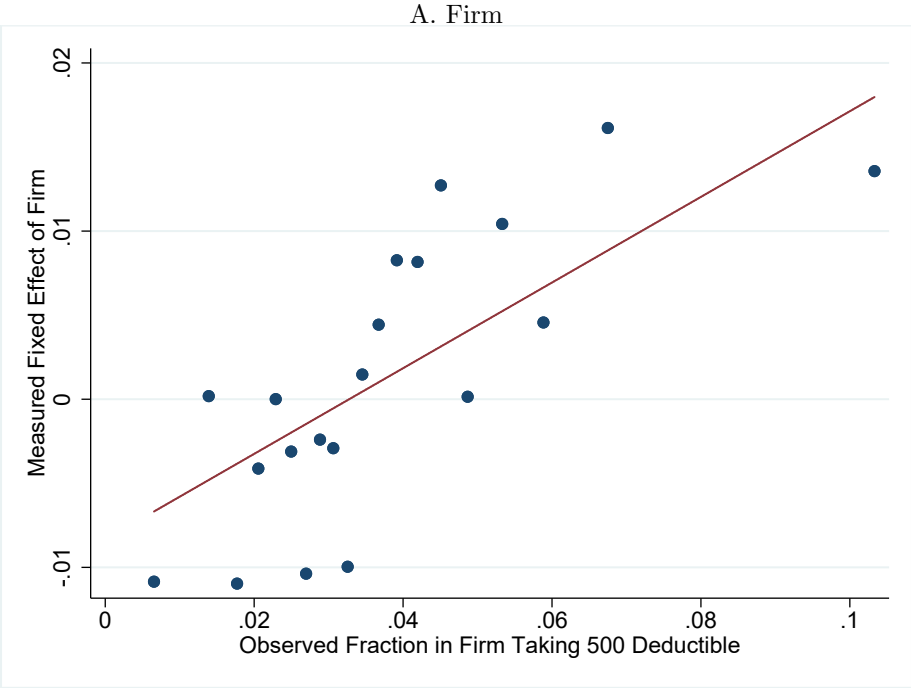
Notes: These figures show binned scatter plots of the relationship between the predicted probability of having costs below 375 EUR (staying under the voluntary deductible range) and the take-up of the voluntary 500 EUR extra deductible, by income group, education level, six selected fields of study and six selected sectors. For these figures, 6 fields of study and 6 sectors were chosen for exposition purposes. Refer to Table 6 and Table 8 for an overview of the take-up of the 500 deductible in all fields and sectors, respectively.

Figure 5: Take-up of 500 Voluntary Deductible as Function of Predicted Health Costs



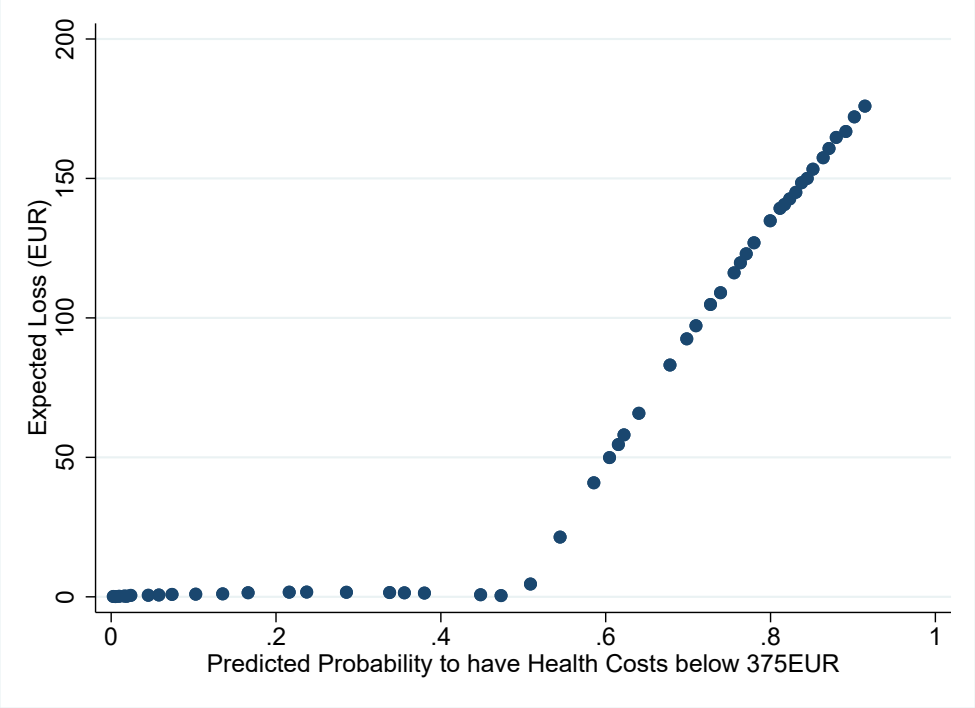
Notes: These figures plot the relationshop between household gross income and the take-up of the 500 EUR extra deductible. Panel A plots take-up of 500 deductible by household income percentile. Panel B plots the residuals of an OLS of take-up of 500 EUR extra deductible on risk probability, 5 level of education dummies, 4 age dummies, gender, partner, children.

Figure 6: Firm Fixed Effects and Deductible Take-Up



Notes: This figure shows a binned scatterplot of the relationship between the firm fixed effect of a regression of the take-up of the 500 EUR extra deductible, as explained in Section V.D, and the observed fraction of the employees in the firm taking a 500 EUR extra deductible. As the firm fixed effect is measured by taking into account individuals who move between firms, this figure suggests a positive causal impact of firms with high deductible take-up on the individual's likelihood of electing a higher deductible.

Figure 7: Expected Loss and Health Cost Probability



Notes: This figure is a binned scatterplot of the relationship between the predicted probability of health costs below 375 EUR and the expected loss due to over- or underinsurance. For individuals with a predicted probability of low costs below 0.5, the expected losses due to underinsurance are very small (on average close to zero), as a very low fraction of people underinsures by taking the 500 EUR extra deductible. For individuals with a predicted probability of low costs above 0.5, expected losses due to overinsurance increase with this probability, and reach almost 170 EUR for people with a very high chance (0.9+) of low costs, as most people leave money on the table by overinsuring for costs that happen with a very low probability.

Figure 8: Predicted Deductible Choice

To be added (under disclosure review).

Notes: This figure shows the distribution of predicted 500 EUR extra deductible take-up rate. Panel A shows the predicted 500 EUR deductible take-up with health effects, while Panel B shows the take-up without the health effects.

Figure 9: Distribution of Choice Quality and Expected Loss

To be added (under disclosure review).

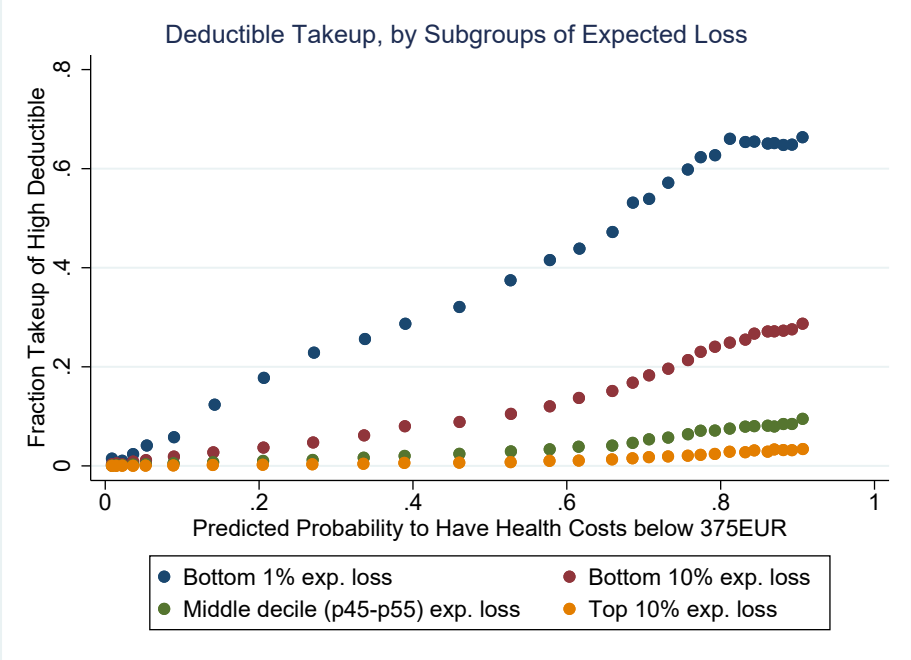
Notes: Panel A shows the distribution of the expected losses due to non-optimal deductible choice decision, controlling for health effects. Panel B shows the distribution of the probability to make the right choice, controlling for health effects.

Figure 10: Choice Quality by Health Risk

To be added (under disclosure review).

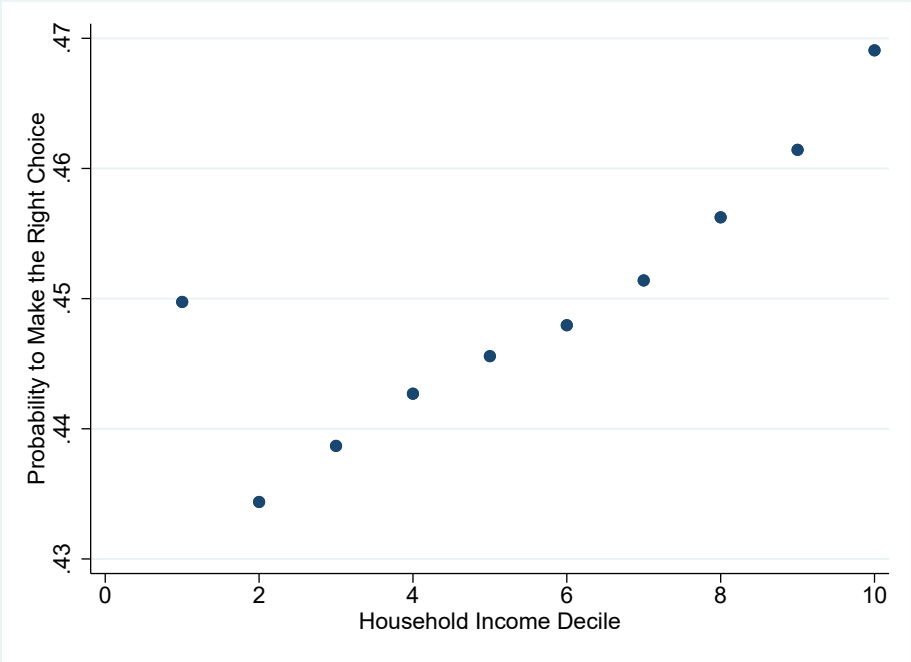
Notes: This figure shows the relationship between the probability to make the right choice and the take-up rate of the 500 EUR extra deductible, controlling for health effects. It is a 8000 random subsample scatter, and hence shows that there is variance in choice quality, even within predicted take-up rates.

Figure 11: Deductible Choice by Subgroups of Expected Loss



Notes: This figure shows a binned scatter plot of the relationship between the predicted probability of having costs below 375 EUR (staying under the voluntary deductible range) and the take-up of the voluntary 500 EUR deductible for four selected subgroups that differ in their expected loss. The bottom 1% expected loss group comes close to a rational consumer, with high take-up of the deductible for low expected costs. The top 10% expected loss group makes the expected loss almost entirely due to overinsurance.

Figure 12: Choice Quality by Household Income Decile



Notes: This figure shows the probability to make the "right" choice of the extra deductible by household income decile. The "right" choice is defined as the choice a rational consumer would make, as explained in Section III. The rational consumer would take the 500 EUR extra deductible if he expects his cost to be below 375 EUR with a probability larger than 0.5. Otherwise, the rational consumer would not take an extra deductible.

Tables

Table 1: Distribution of Deductible Choices

Deductible Type	
Default Deductible (375 EUR)	91.20%
Extra Deductible (+100 to +500 EUR)	8.80%
Breakdown of Extra Deductible Choices	
+100 EUR	11.00%
+200 EUR	10.70%
+300 EUR	6.00%
+400 EUR	1.70%
+500 EUR	70.60%

Notes: This table shows the breakdown of deductible choices in 2015. A large majority (91.2%) sticks to the default 375 EUR deductible. Of the 8.8% individuals that take an extra deductible, most individuals take the 500 EUR extra deductible.

Table 2: Summary Statistics

	Mean	p10	p50	p90
Demographics				
Age	48	23	48	73
Man	48.75%			
Has Children	31.74%			
Partner in Household	63.00%			
Household Financial Status				
Household Gross Income	68299	19440	56090	127281
Household Net Worth	157666	-34674	27926	389813
Mortgage Debt	53.93%			
Other Debt	34.24%			
Savings > 2000 EUR	80.39%			
Education Level				
Unknown	45.72%			
Less than High School	14.00%			
High School	23.87%			
College	10.75%			
Further Studies	5.66%			
Employment Status				
Not Working	44.66%			
Employee	44.97%			
Self-employed or Director	10.36%			
Take-up Deductible in Environment				
Mother Take-up 500	2.89%			
Father Take-up 500	4.04%			
Fraction Take-up 500 in Firm	8.40%	3.48%	7.69%	14.29%
Fraction Take-up 500 in Postcode	6.37%	0.00%	5.00%	14.91%
Total Number of Observations	11606676			

Notes: This table shows summary statistics of the full sample in 2015, excluding the random sample of 1m observations that were used to train the prediction algorithm.

Table 3: Distribution of Annual Health Care Costs

	Mean	p10	p50	p90	p99
All Insurable Care	2 156	0	316	4 777	29 439
Hospital Costs	1 352	0	89	2 836	20 550
Pharmaceuticals	301	0	52	735	3 146
Mental Care	235	0	0	0	4 498
Aids and Tools	101	0	0	125	2 196
Geriatric Care	48	0	0	0	0
Health Care Transport	39	0	0	0	969
Paramedical Care and Physiotherapy	31	0	0	0	1 064
Health Care Received Abroad	15	0	0	0	88
Other Care	33	0	0	74	435
Always Insured Care	197	75	119	292	2 354
Family Doctor Care	154	75	118	271	593
Maternal Care	43	0	0	0	2 162
All Care	2 353	85	476	5 310	29 788

Notes: This table shows the distribution of health expenditures by subcategory, for the full sample in 2014. Expenditures are divided into expenditures subject to cost sharing (to which the deductible applies) versus expenditures that are not subject to cost sharing. All values are in EUR.

Table 4: Deductible Take-Up Regression Main

	(1) OLS	(2) OLS	(3) Probit	(4) Probit
Health Risk				
Dummy Pred. Costs < 375	0.060***	0.045***	0.065***	0.065***
Income				
2nd Income Quartile	-0.001	0.002***	0.003***	0.012***
3rd Income Quartile	0.007***	0.005***	0.009***	0.021***
4th Income Quartile	0.021***	0.011***	0.016***	0.031***
Education Level				
High School	0.011***	0.003***	0.009***	0.009***
Bachelor's Degree	0.045***	0.009***	0.026***	0.019***
Master's Degree	0.069***	0.014***	0.036***	0.024***
Further Studies	0.083***	0.015**	0.039***	0.022***
Demographics				
36 to 50 y.o.	-0.020***	-0.009***	-0.011***	-0.018***
51 to 65 y.o.	-0.020***	-0.009***	-0.011***	-0.019***
65+ y.o.	-0.021***	-0.012***	-0.030***	-0.037***
Gender (Male)	0.022***	0.001*	0.014***	0.005***
Has Partner	0.003***	0.001***	0.003***	0.005***
Has Children	-0.013***	-0.001*	-0.007***	-0.004*
Self-employed	0.024***	0.012***	0.015***	0.023***
Interaction Terms				
2nd Income Quartile x Dummy Pred. Costs < 375		-0.002		-0.012***
3rd Income Quartile x Dummy Pred. Costs < 375		0.005***		-0.014***
4th Income Quartile x Dummy Pred. Costs < 375		0.015***		-0.017***
High School x Dummy Pred. Costs < 375		0.013***		0.001
Bachelor's x Dummy Pred. Costs < 375		0.048***		0.008***
Master's x Dummy Pred. Costs < 375		0.072***		0.013***
Further Studies x Dummy Pred. Costs < 375		0.088***		0.018***
36 to 50 y.o. x Dummy Pred. Costs < 375		-0.012***		0.008***
51 to 65 y.o. x Dummy Pred. Costs < 375		-0.011***		0.009***
65+ y.o. x Dummy Pred. Costs < 375		-0.026***		0.010***
Male x Dummy Pred. Costs < 375		0.030***		0.010***
Has Partner x Dummy Pred. Costs < 375		0.004***		-0.002
Has Children x Dummy Pred. Costs < 375		-0.015***		-0.003
Self-employed x Dummy Pred. Costs < 375		0.016***		-0.009***
Constant	0.009***	0.013***		
Observations	798,430	798,430	798,430	798,430
R-squared	0.043	0.046		

Notes: This table plots coefficients from our regressions studying deductible choice, as explained in Section V. Columns 1 and 2 report coefficients from a linear probability model, while columns 3 and 4 report marginal effects (at mean values) from a probit model. The dependent variable in all specifications is a dummy that takes value of 1 when the individual takes up the voluntary 500 EUR extra deductible. The prob. costs < 375 EUR variable is obtained from our prediction algorithm. The additional variables in specification (2) and (4) are interactions with prob. costs < 375 EUR. The reference groups (groups not entering in the regression as they are perfectly collinear with the other dummies) are: 1st income quartile, education lower than high school or unknown, and age between 18 and 35. *** p<0.01, ** p<0.05, * p<0.1 with robust standard errors.

Table 5: Deductible Take-Up Regression and Wealth, Debt and Liquidity

	(1) OLS	(2) OLS	(3) Probit	(4) Probit
Health Risk				
Dummy Pred. Costs < 375	0.058***	0.039***	0.062***	0.060***
Debt and Liquidity				
Savings > 2000 EUR	0.012***	0.002***	0.011***	0.005***
Mortgage Debt	0.005***	0.003***	0.005***	0.010***
Other Debt	-0.008***	0.000	-0.005***	0.001
Wealth				
2nd Net Worth Quartile	0.001	0.001	-0.002***	-0.003
3rd Net Worth Quartile	0.011***	0.003***	0.007***	0.007***
4th Net Worth Quartile	0.026***	0.007***	0.019***	0.019***
Income				
2nd Income Quartile	-0.009***	-0.000	-0.004***	0.004**
3rd Income Quartile	-0.006***	0.001**	-0.003***	0.008***
4th Income Quartile	0.003***	0.006***	0.001	0.014***
Education Level				
High School	0.011***	0.003***	0.009***	0.009***
Bachelor's Degree	0.044***	0.009***	0.025***	0.018***
Master's Degree	0.068***	0.014***	0.034***	0.023***
Further Studies	0.083***	0.015**	0.039***	0.022***
Demographics				
36 to 50 y.o.	-0.022***	-0.009***	-0.013***	-0.019***
51 to 65 y.o.	-0.028***	-0.011***	-0.018***	-0.024***
65+ y.o.	-0.038***	-0.015***	-0.043***	-0.046***
Gender (Male)	0.022***	0.001*	0.014***	0.005***
Has Partner	0.003***	0.001**	0.004***	0.003**
Has Children	-0.011***	-0.001	-0.006***	-0.003
Self-employed	0.019***	0.010***	0.010***	0.018***
Interaction Terms				
Savings > 2000 EUR x Dummy Pred. Costs < 375		0.018***		0.007***
Mortgage Debt x Dummy Pred. Costs < 375		0.004***		-0.006***
Other Debt x Dummy Pred. Costs < 375		-0.011***		-0.007***
2nd Net Worth Quartile x Dummy Pred. Costs < 375		-0.003**		0.000
3rd Net Worth Quartile x Dummy Pred. Costs < 375		0.012***		-0.000
4th Net Worth Quartile x Dummy Pred. Costs < 375		0.033***		-0.000
2nd Income Quartile x Dummy Pred. Costs < 375		-0.010***		-0.009***
3rd Income Quartile x Dummy Pred. Costs < 375		-0.011***		-0.012***
4th Income Quartile x Dummy Pred. Costs < 375		-0.008***		-0.015***
High School x Dummy Pred. Costs < 375		0.014***		0.001
Bachelor's x Dummy Pred. Costs < 375		0.047***		0.008***
Master's x Dummy Pred. Costs < 375		0.070***		0.012***

Further Studies x Dummy Pred. Costs < 375		0.089***		0.018***
36 to 50 y.o. x Dummy Pred. Costs < 375		-0.016***		0.007***
51 to 65 y.o. x Dummy Pred. Costs < 375		-0.023***		0.008***
65+ y.o. x Dummy Pred. Costs < 375		-0.049***		0.007***
Male x Dummy Pred. Costs < 375		0.030***		0.010***
Has Partner x Dummy Pred. Costs < 375		0.005***		0.000
Has Children x Dummy Pred. Costs < 375		-0.011***		-0.002
Self-employed x Dummy Pred. Costs < 375		0.010***		-0.008***
Constant	0.007***	0.012***		
Observations	796,524	796,524	796,524	796,524
R-squared	0.046	0.050		

Notes: The notes from Table 4 apply. Additional variables are a dummy for liquidity (household savings > 2000), a dummy for having household mortgage debt and other household debt, household net worth quartiles, as well as their interactions with the health risk. *** p<0.01, ** p<0.05, * p<0.1 with robust standard errors.

Table 6: Deductible Choice and Field of Study

Education Field	(A) Predicted Costs < 375	(B) Taking 500 Deductible	(C) Deductible 500 | Predicted Costs < 375	(D) Taking 100-400 Deductible	Observations
Statistics	**87%**	**29%**	**34%**	**3%**	**593**
Mathematics	85%	21%	27%	4%	2551
Physics	91%	21%	26%	4%	5888
Architecture and Town Planning	88%	18%	21%	5%	17960
Physical Science	82%	18%	22%	4%	484
Earth Science	88%	18%	21%	5%	2287
Philosophy and Ethics	**82%**	**17%**	**21%**	**4%**	**2603**
Medicine	83%	17%	20%	5%	33834
Chemistry	87%	16%	20%	5%	3215
Biology and Biochemistry	83%	16%	20%	5%	7526
Science, Mathematics and Computing	85%	16%	19%	4%	2107
Computer Science	87%	15%	18%	5%	83778
Environmental Protection	86%	15%	18%	5%	6040
Political Science and Civics	85%	15%	18%	6%	4292
Design	85%	15%	18%	4%	12823
Sociology and Cultural Studies	82%	14%	18%	5%	19334
Mining and Extraction	91%	14%	17%	4%	814
Economics	84%	14%	17%	4%	60546
Humanities and Arts	84%	14%	18%	3%	462
Dental Studies	76%	14%	18%	3%	9339
History and Archaeology	82%	13%	16%	5%	17315
Business and Administration	82%	13%	16%	4%	78620
Pharmacy	73%	13%	17%	3%	10232
Health	79%	13%	16%	4%	18337
Environmental Protection Technology	84%	13%	15%	4%	2301
Medical Diagnostic and Treatment Technology	81%	13%	16%	4%	21236
Religion	80%	13%	17%	4%	3759
Law	80%	13%	16%	4%	60346
Psychology	77%	12%	16%	4%	31745
Management and Administration	81%	12%	16%	4%	147286
Engineering and Engineering Trades	87%	12%	15%	4%	26090
Forestry	86%	12%	14%	4%	2608
Therapy and Rehabilitation	78%	12%	15%	4%	31029
Finance, Banking, Insurance	80%	12%	15%	4%	14523
Social and Behavioural Science	79%	12%	15%	4%	43533
Health and Welfare	80%				59
Fisheries	94%	12%	15%		93
Journalism and Reporting	80%	12%	14%	5%	9014
Training for Teachers with Subject Specialisation	79%	11%	14%	5%	29249
Education Science	75%	11%	14%	4%	27346
Accounting and Taxation	**78%**	**11%**	**14%**	**3%**	**65663**
Agriculture, Forestry and Fishery	81%	10%	13%	4%	12320
Marketing and Advertising	**80%**	**10%**	**13%**	**4%**	**30720**
Chemical and Process	85%	10%	12%	3%	23420

Arts	80%	10%	13%	4%	5383
Electronics and Automation	86%	10%	12%	4%	10589
Music and Performing Arts	81%	10%	12%	4%	11280
Training for Teachers of Vocational Subjects	81%	10%	12%	4%	31289
Fine Arts	82%	10%	12%	4%	17904
Humanities	76%	10%	12%	4%	798
Library, Information, Archive	78%	9%	12%	5%	2894
Travel, Tourism and Leisure	77%	9%	12%	4%	42647
Electricity and Energy	88%	9%	11%	3%	61444
Veterinary	75%	9%	12%	3%	6462
Mother Tongue	74%	9%	12%	4%	3832
Audio-visual Techniques and Media Production	83%	9%	10%	4%	28743
Building and Civil Engineering	86%	9%	10%	3%	110817
Life Science	79%	9%	11%	4%	1325
Crop and Livestock Production	79%	9%	11%	4%	32497
Mechanics and Metal Work	85%	9%	10%	3%	70897
Wholesale and Retail Sales	79%	8%	11%	3%	155053
Foreign Languages	74%	8%	11%	3%	17585
Motor Vehicles, Ships and Aircraft	87%	8%	10%	3%	39471
Training for Teachers at Basic Levels	75%	8%	10%	4%	72790
Materials (Wood, Paper, Plastic, Glass)	86%	8%	9%	3%	5653
Sports	83%	8%	10%	3%	13626
Teacher Training and Education Science	74%	8%	10%	4%	24559
Military and Defence	81%	7%	9%	3%	1516
Transport Services	83%	7%	9%	3%	60138
Food Processing	78%	7%	9%	3%	21128
Natural Environments and Wildlife	86%				80
Hotel, Restaurant and Catering	77%	6%	8%	3%	64169
Basic / Broad, General Programmes	72%	6%	9%	2%	501278
Social Work and Counselling	70%	6%	8%	3%	149817
Personal Skills	68%	6%	8%	3%	1961
Textiles, Clothes, Footwear, Leather	70%	5%	7%	3%	15138
Horticulture	80%	5%	6%	3%	12190
General Programmes	71%	5%	7%	2%	8101
Nursing and Caring	66%	5%	7%	2%	126843
Domestic Services	66%	5%	7%	2%	84328
Secretarial and Office Work	65%	5%	7%	2%	47200
Protection of Persons and Property	**78%**	**4%**	**6%**	**2%**	**49349**
Child Care and Youth Services	66%	4%	6%	2%	7489
Computer Use	65%	4%	6%	2%	3051
Hair and Beauty Services	**65%**	**4%**	**5%**	**2%**	**35157**
Occupational Health and Safety	75%				108
Training for Pre-school Teachers	62%				77
Literacy and Numeracy	62%	2%	4%	2%	905

Notes: This table shows by level of field of study: (A) the fraction of individuals who have predicted costs < 375 EUR, (B) the fraction of individuals who take-up a 500 EUR extra deductible, (C) the fraction of individuals who take-up a 500 EUR extra deductible, conditional on having predicted health costs < 375 EUR, and (D) the fraction of individuals who take-up an intermediary voluntary deductible of 100, 200, 300 or 400 EUR. To limit the danger of disclosure, cells with underlying $n < 10$ are emptied.

Table 7: Deductible Take-Up Regression and Field of Study

	(1) OLS	(2) OLS	(3) Probit	(4) Probit
Health Risk				
Dummy Pred. Costs < 375	0.068***	0.034***	0.089***	0.079***
Education Field				
Statistics	0.123**	0.099	0.050**	0.097
Philosophy	0.048**	0.014	0.028***	0.031
Accounting and Taxation	0.002	0.000	0.001	0.002
Marketing and Advertising	-0.005	-0.000	-0.003	0.000
Hair and Beauty	-0.014***	-0.002	-0.016***	-0.006
Protection of Persons	-0.030***	-0.001	-0.031***	-0.005
Income				
2nd Income Quartile	-0.005***	0.001*	-0.003**	0.008**
3rd Income Quartile	0.002**	0.005***	0.006***	0.018***
4th Income Quartile	0.014***	0.011***	0.013***	0.032***
Education Level				
High School	0.022***	0.004***	0.031***	0.020***
Bachelor's Degree	0.057***	0.011***	0.056***	0.036***
Master's Degree	0.081***	0.015***	0.069***	0.043***
Further Studies	0.096***	0.016***	0.076***	0.043***
Demographics				
36 to 50 y.o.	-0.022***	-0.010***	-0.018***	-0.029***
51 to 65 y.o.	-0.022***	-0.009***	-0.019***	-0.026***
65+ y.o.	-0.025***	-0.013***	-0.041***	-0.049***
Gender (Male)	0.027***	0.001	0.021***	0.005
Has Partner	0.002***	0.001	0.002**	0.003
Has Children	-0.012***	-0.000	-0.008***	-0.001
Self-employed	0.020***	0.013***	0.016***	0.030***
Interaction Terms				
Statistics x Dummy Pred. Costs < 375		0.021		-0.051
Philosophy x Dummy Pred. Costs < 375		0.050		-0.003
Accounting and taxation x Dummy Pred. Costs < 375		0.002		-0.000
Marketing and advertising x Dummy Pred. Costs < 375		-0.006		-0.004
Hair and beauty x Dummy Pred. Costs < 375		-0.017***		-0.012
Protection of persons x Dummy Pred. Costs < 375		-0.041***		-0.030***
2nd Income Quartile x Dummy Pred. Costs < 375		-0.008***		-0.012***
3rd Income Quartile x Dummy Pred. Costs < 375		-0.001		-0.014***
4th Income Quartile x Dummy Pred. Costs < 375		0.004*		-0.022***
High School x Dummy Pred. Costs < 375		0.032***		0.012***
Bachelor's x Dummy Pred. Costs < 375		0.068***		0.023***
Master's x Dummy Pred. Costs < 375		0.093***		0.029***
Further Studies x Dummy Pred. Costs < 375		0.111***		0.037***

36 to 50 y.o. x Dummy Pred. Costs < 375		-0.015***		0.012***
51 to 65 y.o. x Dummy Pred. Costs < 375		-0.017***		0.008**
65+ y.o. x Dummy Pred. Costs < 375		-0.030***		0.012**
Male x Dummy Pred. Costs < 375		0.034***		0.018***
Has Partner x Dummy Pred. Costs < 375		0.001		-0.002
Has Children x Dummy Pred. Costs < 375		-0.014***		-0.008**
Self-employed x Dummy Pred. Costs < 375		0.009***		-0.015***
Constant	-0.004***	0.013***		
Observations	430,765	430,765	430,765	430,765
R-squared	0.038	0.042		

Notes: The notes from Table 4 apply. Additional variables are dummies for six selected fields of study, as well as their interactions with the health risk. The reference category for field of study is all other fields of study. *** p<0.01, ** p<0.05, * p<0.1 with robust standard errors.

Table 8: Deductible Choice and Sector

Sector of Employment	(A) Predicted Costs < 375	(B) Taking 500 Deductible	(C) Deductible 500 | Predicted Costs < 375	(D) Taking 100-400 Deductible	Observations
Business Services II	**84%**	**13%**	**16%**	**5%**	**376695**
Insurance and Health Insurance Firms	**79%**	**12%**	**15%**	**3%**	**54916**
Business Services I	82%	12%	15%	4%	85091
Dairy Industry	82%	12%	14%	4%	16271
Banks	81%	10%	12%	4%	86285
Other Passenger Transport Land and Air	79%	10%	13%	4%	6425
Business Services III	79%	10%	13%	4%	391323
Agriculture	85%	10%	11%	3%	87657
Stoneware	83%	9%	11%	3%	383
Publishers	79%	9%	11%	5%	22229
Cultural Institutions	80%	9%	11%	4%	51310
Telecommunications	81%	9%	12%	3%	30613
Government, Education and Science	75%	9%	12%	4%	477053
Food Industry	80%	9%	11%	4%	49854
Catering Industry I	84%	9%	10%	3%	281784
Tobacco Processing Industry	76%	9%	11%	5%	2198
Wholesale I	82%	8%	11%	3%	150898
Wholesale II	81%	8%	10%	4%	227459
Government, Police and Judiciary	74%	8%	11%	4%	176143
Wholesale of Wood	82%	8%	10%	3%	5541
Electronic Industry	81%	8%	13%	3%	38791
Carpentry	83%	8%	9%	3%	6333
Furniture and Organ Building	83%	8%	9%	3%	12000
Rail Construction	78%	8%	11%	3%	7593
Railway Transport	74%	8%	10%	3%	15147
Sugar Processing Industry	78%	7%	10%	3%	10508
Chain Stores	80%	7%	9%	3%	229871
Retail	**79%**	**7%**	**9%**	**3%**	**314586**
Lending Industry	81%	7%	9%	3%	240139
Other Branches of Business	79%	7%	9%	3%	38766
Postal Transport	72%	7%	10%	3%	42996
Metal Industry	80%	7%	10%	3%	119469
Construction	**83%**	**7%**	**9%**	**3%**	**99088**
Merchant	89%	7%	8%	3%	5452
Mortar	72%	7%	9%	3%	1275
Flight Transport	77%	7%	9%	3%	29198
Bakeries	79%	7%	9%	3%	27663
Metal and Technical Industry	82%	7%	8%	3%	330795
Port Companies	82%	7%	9%	3%	91645
Chemical Industry	79%	7%	9%	3%	61429
General Industry	81%	7%	9%	2%	61092
Stone, Cement, Glass and Ceramic Industry	77%	7%	9%	3%	16089

Butchers Other	80%	7%	8%	3%	16075
Health, Mental and Social Industry	71%	7%	9%	3%	1218825
Printing Industry	80%	7%	8%	3%	21771
Textiles Industry	77%	7%	9%	2%	5710
Inland Shipping	83%	7%	8%	3%	6379
Private Bus Transport	70%	6%	9%	2%	3065
Government, Local Government	70%	6%	9%	3%	195152
Butchers	79%	6%	8%	3%	11742
Wood, Brush and Packaging Industry	82%	6%	8%	3%	5797
Other Goods Transport Land and Air	80%	6%	8%	3%	127703
Government, Defense	82%	6%	11%	3%	26065
Government, Public Utilities	**77%**	**6%**	**7%**	**2%**	**15353**
Public Transport	65%	5%	8%	2%	18935
Security	75%	5%	7%	2%	24907
Plastering	85%	5%	6%	3%	2821
Taxi and Ambulance	65%	5%	8%	2%	17868
Catering Industry II	70%	5%	7%	2%	17865
Painting Industry	81%	5%	6%	3%	11590
Port Classifiers	79%	5%	6%	3%	2821
Fishing	81%	4%	6%	3%	825
Work and Integration	64%	4%	6%	2%	8915
Dredging Industry	85%	4%	9%	2%	6050
Government, Other Institutions	60%	4%	7%	2%	146892
Roofing	82%	4%	5%	2%	2919
Cleaning	**70%**	**3%**	**5%**	**2%**	**94172**

Notes: The interpretation for this table is similar as the one for table 6. Individuals that are not working, or working in a firm in an unknown sector, are excluded.

Table 9: Deductible Take-Up Regression and Sector

	(1) OLS	(2) OLS	(3) Probit	(4) Probit
Health Risk				
Dummy Pred. Costs < 375	0.068***	0.052***	0.094***	0.082***
Professional Sector				
Business services	0.027***	0.004	0.017***	0.010*
Insurance	0.035***	0.008	0.025***	0.018
Retail	-0.001	0.001	-0.001	0.004
Construction	-0.006*	-0.001	-0.005*	-0.002
Cleaning	-0.019***	-0.003	-0.034***	-0.015
Public utilities	-0.015*	0.006	-0.014*	0.015
Income				
2nd Income Quartile	-0.009***	-0.001	-0.009***	-0.003
3rd Income Quartile	-0.001	0.001	-0.000	0.004
4th Income Quartile	0.012***	0.007***	0.012***	0.020***
Education Level				
High School	0.009***	0.002*	0.011***	0.007**
Bachelor's Degree	0.043***	0.008***	0.037***	0.022***
Master's Degree	0.069***	0.014***	0.052***	0.032***
Further Studies	0.091***	0.016**	0.063***	0.035***
Demographics				
36 to 50 y.o.	-0.020***	-0.009***	-0.017***	-0.025***
51 to 65 y.o.	-0.020***	-0.010***	-0.018***	-0.029***
65+ y.o.	-0.027***	-0.014***	-0.041***	-0.049***
Gender (Male)	0.026***	0.001	0.021***	0.003
Has Partner	0.003***	0.001	0.003***	0.004
Has Children	-0.012***	-0.002	-0.009***	-0.004
Self-employed	0.026***	0.014***	0.022***	0.035***
Interaction Terms				
Business services x Dummy Pred. Costs < 375		0.026***		0.007
Insurance x Dummy Pred. Costs < 375		0.034***		0.008
Retail x Dummy Pred. Costs < 375		-0.002		-0.005
Construction x Dummy Pred. Costs < 375		-0.006		-0.004
Cleaning x Dummy Pred. Costs < 375		-0.026***		-0.021*
Public utilities x Dummy Pred. Costs < 375		-0.028**		-0.032
2nd Income Quartile x Dummy Pred. Costs < 375		-0.011***		-0.007
3rd Income Quartile x Dummy Pred. Costs < 375		-0.002		-0.004
4th Income Quartile x Dummy Pred. Costs < 375		0.007***		-0.009*
High School x Dummy Pred. Costs < 375		0.011***		0.005
Bachelor's x Dummy Pred. Costs < 375		0.045***		0.016***
Master's x Dummy Pred. Costs < 375		0.069***		0.022***
Further Studies x Dummy Pred. Costs < 375		0.091***		0.030**

36 to 50 y.o. x Dummy Pred. Costs < 375		-0.013***		0.009***
51 to 65 y.o. x Dummy Pred. Costs < 375		-0.012***		0.012***
65+ y.o. x Dummy Pred. Costs < 375		-0.028***		0.010
Male x Dummy Pred. Costs < 375		0.030***		0.019***
Has Partner x Dummy Pred. Costs < 375		0.002		-0.002
Has Children x Dummy Pred. Costs < 375		-0.012***		-0.005
Self-employed x Dummy Pred. Costs < 375		0.015***		-0.015***
Constant	0.008***	0.018***		
Observations	443,893	443,893	443,893	443,893
R-squared	0.030	0.032		

Notes: The notes from Table 4 apply. Additional variables are dummies for specific sectors, as well as their interactions. The reference category for sector is all other sectors. *** p<0.01, ** p<0.05, * p<0.1 with robust standard errors.

Table 10: Deductible Take-Up Regression and Environment

	(1) OLS	(2) OLS	(3) Probit	(4) Probit
Health Risk				
Dummy Pred. Costs < 375	0.068***	-0.018***	0.100***	0.069***
Environment variables				
Share of Colleagues with 500 Deductible	0.288***	0.060***	0.232***	0.158***
Share of People in Postcode with 500 Deductible	0.409***	0.152***	0.334***	0.349***
Father Has Taken 500 Deductible	0.191***	0.094***	0.094***	0.093***
Mother Has Taken 500 Deductible	0.255***	0.120***	0.113***	0.098***
Income				
2nd Income Quartile	-0.002	0.000	-0.001	0.002
3rd Income Quartile	0.002	0.000	0.004*	0.001
4th Income Quartile	0.008***	0.006*	0.008***	0.013
Education Level				
High School	0.007***	0.001	0.010***	0.005
Bachelor's Degree	0.033***	0.005**	0.033***	0.020***
Master's Degree	0.056***	0.004	0.048***	0.018**
Further Studies	0.087***	0.009	0.067***	0.029
Demographics				
36 to 50 y.o.	-0.003**	-0.004**	-0.003**	-0.013**
51 to 65 y.o.	0.009***	-0.003	0.008***	-0.012
65+ y.o.	-0.053***	-0.012***		
Gender (Male)	0.029***	0.001	0.025***	0.003
Has Partner	0.007***	0.000	0.008***	0.006
Has Children	-0.007***	-0.001	-0.006***	-0.004
Self-employed	0.019***	0.005	0.016***	0.014
Interaction Terms				
Share of Colleagues w/ 500 Ded. x Dummy Pred. Costs < 375		0.267***		0.080*
Share in Postcode w/ 500 Ded. x Dummy Pred. Costs < 375		0.309***		-0.015
Father has Taken 500 Ded. x Dummy Pred. Costs < 375		0.109***		0.002
Mother has Taken 500 Ded. x Dummy Pred. Costs < 375		0.149***		0.017
2nd Income Quartile x Dummy Pred. Costs < 375		-0.001		-0.003
3rd Income Quartile x Dummy Pred. Costs < 375		0.005		0.003
4th Income Quartile x Dummy Pred. Costs < 375		0.005		-0.006
High School x Dummy Pred. Costs < 375		0.009***		0.006
Bachelor's x Dummy Pred. Costs < 375		0.034***		0.014**
Master's x Dummy Pred. Costs < 375		0.062***		0.032***
Further Studies x Dummy Pred. Costs < 375		0.093***		0.041*
36 to 50 y.o. x Dummy Pred. Costs < 375		0.001		0.011**
51 to 65 y.o. x Dummy Pred. Costs < 375		0.015***		0.023**
65+ y.o. x Dummy Pred. Costs < 375		-0.051***		
Male x Dummy Pred. Costs < 375		0.033***		0.024***

Has Partner x Dummy Pred. Costs < 375		0.007***		0.001
Has Children x Dummy Pred. Costs < 375		-0.006		-0.002
Self-employed x Dummy Pred. Costs < 375		0.018**		0.002
Constant	-0.069***	-0.000		
Observations	194,663	194,663	194,654	194,654
R-squared	0.114	0.119		

Notes: The notes from Table 4 apply. Additional variables are the fraction of individuals taking up an extra 500 EUR deductible in firm and neighbourhood, and dummies for whether the father or mother is taking up an extra 500 EUR deductible. Note that the shares are calculated excluding the individual for which the share is calculated (ie. the person's take-up is excluded from both numerator and denominator), and shares are calculated only if there are more than 10 individuals that firm or neighbourhood. *** p<0.01, ** p<0.05, * p<0.1 with robust standard errors.

Table 11: Health Shocks on Deductible Take-up

	(1)	(2)	(3)	(4)	(5)
	No FE	Individual FE	First Difference	First Difference	First Difference
Health	0.0875***	0.0453***	0.0284***		
Positive △ Health				0.0156***	
Negative △ Health				-0.0467***	
△ Health > +2 Deciles					0.0176***
△ Health = +2 Deciles					0.0173***
△ Health = +1 Decile					0.0136***
△ Health = -1 Decile					-0.0468***
△ Health = -2 Deciles					-0.0341***
△ Health < -2 Deciles					-0.0518***
Constant	-0.0158***	0.00310***	0.00465***	0.00580***	0.00567***
Observations	3,499,585	3,499,585	2,534,188	2,528,395	2,534,182
Number of persons	799,906	799,906	741,394	741,394	741,394

Notes: This table presents the result of an OLS regressions of take-up of the 500 EUR extra deductible on health risk and changes in health risk. In column (1), take-up of the high deductible is regressed on health decile level; i.e. the probability to have health costs lower than 375 EUR. Column (2) adds individual fixed effects. Column (3) regresses the first difference of deductible take-up on the first difference of health risk. Column (4) splits the first difference in two distinct variables, one containing only positive shocks, the other only negative shocks. Column (6) splits shocks into 3 categories of magnitude on the positive and the negative side. All regressions include year fixed effects. *** p<0.01, ** p<0.05, * p<0.1 with robust standard errors.

Table 12: Health and Income Shocks and Deductible Take-up

	(1)	(2)	(3)
	No Individual FE	Individual FE	First Difference
Health	0.0947***	0.0422***	0.0312***
Income	0.000922***	-8.23e-05	-2.80e-05
Constant	-0.0210***	0.00929***	0.0250***
Observations	2,842,720	2,842,720	1,934,573
Number of persons	799,861	799,861	729,571

Notes: This table presents the result of an OLS regressions of take-up of the 500 EUR extra deductible on health risk deciles and income deciles. In column (1), take-up of the 500 EUR extra deductible is regressed on both variables in decile levels; column (2) adds individual fixed effects. Column (3) regresses instead the first difference of deductible take-up on the first difference of income deciles, and the first difference of health deciles. The income relates to the gross household income, and health risk is the probability to have health costs lower than 375 EUR. All regressions include year fixed effects. *** p<0.01, ** p<0.05, * p<0.1 with robust standard errors.

Table 13: Best and Worst Decision Makers

	Best Decision Makers (Bottom 5% Expected Loss)	Worst Decision Makers (Top 5% Expected Loss)
Demographics		
Gender (Male)	57%	29%
Age	32	39
Gross Income	90913	50377
Has Children	21.91%	16.21%
Has a Partner	48.86%	31.58%
Education Level		
Less than High School	7.78%	76.84%
High School	39.34%	23.09%
Some College	31.74%	0.04%
College and Further	21.15%	0.03%
Education Field		
Statistics	4.73%	0.00%
Philosophy	11.06%	0.00%
Tax and Administration	37.84%	0.45%
Marketing and Advertising	19.26%	5.11%
Hair and Beauty Services	14.70%	30.38%
Protection of Persons	12.42%	64.06%
Professional Sector		
Business Services	60.01%	8.72%
Insurance	5.99%	0.19%
Retail	25.69%	46.47%
Construction	5.31%	7.87%
Cleaning	2.14%	36.37%
Public Utilities	0.87%	0.38%

Notes: This table presents observable characteristics for the groups that our model considers to be the top 5% and the bottom 5% decision makers. This table is best interpreted by, for each characteristic, comparing the frequency of that characteristic in the best versus the worst decision maker group. For example, 21.15% of best decision makers have a college or more advanced degree, while only 0.03% of the worst decision makers have such a degree.

Table 14: Welfare Impact of Alternative Policies [PRELIMINARY!]

	Not Controlling for Health Effects			Controlling for Health Effects		
Risk-neutral case	Optimal	Mandate 500	Mandate 0	Optimal	Mandate 500	Mandate 0
Unweighted	80.6	20.6	-10.4	70.7	-1.5	0.1
Low Inequality Aversion (e=0.5)	70.8	5.7	-8.7	67.5	-1.4	0.1
High Inequality Aversion (e=1.5)	47.0	-39.8	-4.7	62.3	-1.3	0.1
Sigma = 0,00001	Optimal	Mandate 500	Mandate 0	Optimal	Mandate 500	Mandate 0
Unweighted	80.4	20.4	-10.4	70.6	-1.6	0.1
Low Inequality Aversion (e=0.5)	70.7	5.5	-8.7	67.4	-1.6	0.1
High Inequality Aversion (e=1.5)	46.9	-39.9	-4.7	62.2	-1.5	0.1
Sigma = 0,0001	Optimal	Mandate 500	Mandate 0	Optimal	Mandate 500	Mandate 0
Unweighted	79.1	18.7	-10.2	69.5	-3.2	0.2
Low Inequality Aversion (e=0.5)	69.5	3.9	-8.6	66.3	-3.1	0.2
High Inequality Aversion (e=1.5)	46.0	-41.3	-4.6	61.2	-2.9	0.1
Sigma = 0,001	Optimal	Mandate 500	Mandate 0	Optimal	Mandate 500	Mandate 0
Unweighted	67.2	0.7	-8.8	59.8	-19.2	1.4
Low Inequality Aversion (e=0.5)	58.9	-12.8	-7.3	57.0	-18.4	1.2
High Inequality Aversion (e=1.5)	38.4	-55.3	-3.8	52.3	-17.2	0.9

Notes: This table shows the average welfare impact (in EUR per person) of three alternative policies concerning the extra deductible: optimal (all individuals taking the optimal deductible given their health risk), mandate 500 (all individuals take the 500 EUR extra deductible), and mandate 0 (no one takes the extra deductible). The welfare impact is calculated not controlling for health (left panel), or controlling for health (right panel). The welfare impact is also calculated with equal weights, low inequality aversion or high inequality aversion. Weights y_δ are computed as $y_\delta^{-\epsilon} / \left(\bar{y_\delta^{-\epsilon}}\right)$ for $\epsilon = .5$ and $\epsilon = 1.5$. The sample excludes individuals with a gross income below the social minimum, students, and self-employed people.

A Health Cost Predictions

In this Appendix, we discuss the prediction algorithm that we use to obtain risk probabilities. We also include the case for opting for the binary prediction of health cost risk.

A.A Prediction Algorithm

We use a machine learning algorithm to obtain predictions on the probability that an individual's health costs will be below 375 EUR in any given year. The prediction method we use is a standard and well-known machine learning method, an ensemble learner that consists of a random forest model, gradient boosted regression trees and LASSO. To avoid overfitting, we train and calibrate the prediction algorithm solely on a "training sample" of 1 million individuals. We then use this trained prediction algorithm to obtain predictions for our hold-out sample of 11 million individuals. All the analyses and statistics in the paper are developed use only this hold-out sample.

The prediction method we use follows four steps, which closely resemble the steps used in Einav et al. (2018). First, we follow standard practice in machine learning by tuning key parameters that govern the prediction models by 3-fold cross-validation. Second, we train the three resulting prediction models separately. Third, we combine the three obtained predictions into one using a linear combination that we calibrate in the data. Finally, we calibrate the resulting final ensemble predictions using a linear spline. We describe each of these steps in more detail here.[28]

Parameter Tuning As the three machine learning models that we use have parameters that are choosable by the researcher, we follow standard practice and tune these parameters using 3-fold cross validation. More specifically, we tune the following parameters: minimal node size (mid.node.size), number of variables used at each node (mtry) for the random forest model, learning rate (eta) for the boosted regression trees, and the shrinkage parameter (lambda) for the LASSO.[29] For each of these parameters, we optimize among 5 alternatives. We tune these parameters using 3-fold cross validation, where we are optimizing the area under the receiver operating characteristic curve (AUC).[30] Thus, for each of the parameter values we want to test, the model is trained on 2 folds (subsets of the training sample), and then the performance is measured in the 3rd fold. The parameter values for which the AUC in the hold-out sample is highest for each prediction algorithm are: mtry = 40, min.node.size = 1, eta = 0.3, lambda = 0.001.

[28] As there is some variation in the number and definition of predictors that we have across time, we repeat these four steps for all years of study (2011-2015).

[29] We use the package CARET in R that provides a standardized way to tune parameters. The prediction models we use are RANGER (random forest), XGBLINEAR (boosted regression trees), and GLMNET (LASSO).

[30] This is a common metric used in the machine learning literature to measure the performance of a prediction model.

Estimating the Models Using these tuned parameter values, all models are estimated using a training sample of 800k individuals.

Obtaining Ensemble Predictor We then combine the predictions from the random forest, gradient boosting regression trees, and LASSO into one ensemble prediction. Following Einav et al. (2018), we construct the ensemble predictor to be the linear combination $p_{ensemble} = \hat{\beta}_{rf}\hat{p}_{rf} + \hat{\beta}_{gb}\hat{p}_{gb} + \hat{\beta}_{lasso}\hat{p}_{lasso}$, where $\hat{p}_x$ is the prediction from algorithm x and $\hat{\beta}_x$ is the associated weight.

We obtain estimates for the weights from a constrained linear regression (with no constant and the weights summing to one) of the dummy for having costs below 375 on the three individual predicted probabilities. For this step, we use 100k observations that we did not use in either step of parameter tuning or the estimation on the models. We find associated weights in 2015 that are $\hat{\beta}_{rf} = 0.67$, $\hat{\beta}_{gb} = 0.08$ and $\hat{\beta}_{lasso} = 0.25$.

Calibrating Probabilities Finally, these raw probability predictions are calibrated to the actual observed probabilities. This calibration is estimated using 100k observations that are used in none of the previous steps. From these 100k observations, 100 bins are created and in every bin the mean probability is calibrated to the observed mean probability. All probabilities are then obtained from linear interpolation using those calibrated probabilities.

A.B Results of Prediction

Figure 2 shows the distribution of predicted probabilities using this prediction algorithm on the hold-out sample. It shows that there is a lot over variation in the predicted costs among individuals. Interestingly, there is little mass with a higher than 95% chance of having expected costs below 375. Figure 13 shows the predicted and observed spending probabilities for bins where the individuals have not been used at all in any of the ML estimation or calibration, and shows that the prediction of the spending buckets with our calibration method matches observed spending patterns quite well.

Figure 14 presents the importance of different predictors for the random forest results.[31] Not surprisingly, the most important predictors are different categories of past medical spending, with 2014 values being more important than 2013 values. Spending on drugs and hospital spending are the most important predictors, while age and visits to primary care doctors are also important predictors.

[31] Note that the random forest is only one part of our prediction algoritm, yet is the most important one as it receives the highest weight in the ensemble learner.

A.C Non-Binary Prediction

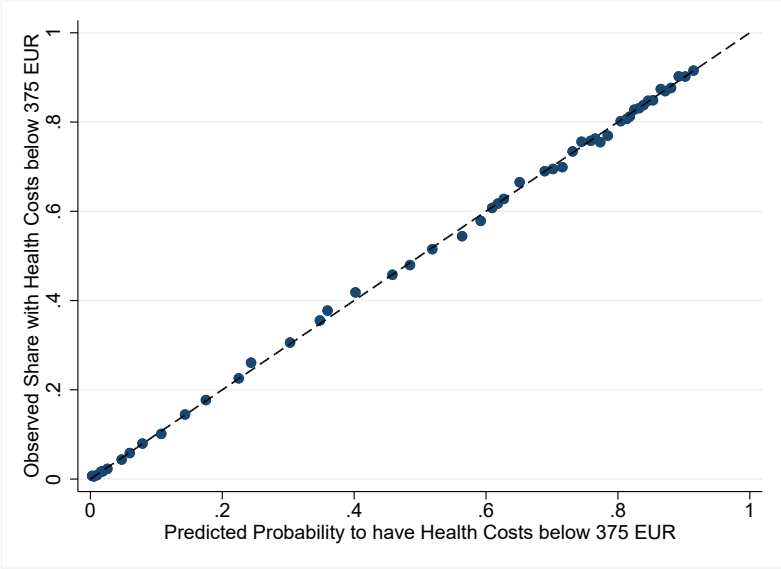
In Section IV, we argue that a useful cost prediction for our application is a binary prediction of having costs lower or higher than 375 EUR, which is the cut-off for the lowest deductible in 2015. This is a simplification, as in fact there are 6 different deductible choices possible, which apply to different brackets with cutoffs at 375, 475, 575, 675, 775 and 875 EUR. In this Appendix section, we strengthen our case that the binary prediction is appropriate, as the probability to be in one of the intermediate deductible categories is negligable.

Panel A of Figure 15 shows the ex-post observed shares within each deductible bracket. Ex post, most individuals fall in either the lowest or highest deductible brackets, yet there is mass on the intermediary deductible ranges.

Predicting with high probability that someone will have costs in those intermediary ranges, however, is difficult. The predicted probabilities in Panel B show that ex-ante, a random forest model trained on an unbalanced sample will only give negligible probability to the intermediate categories.

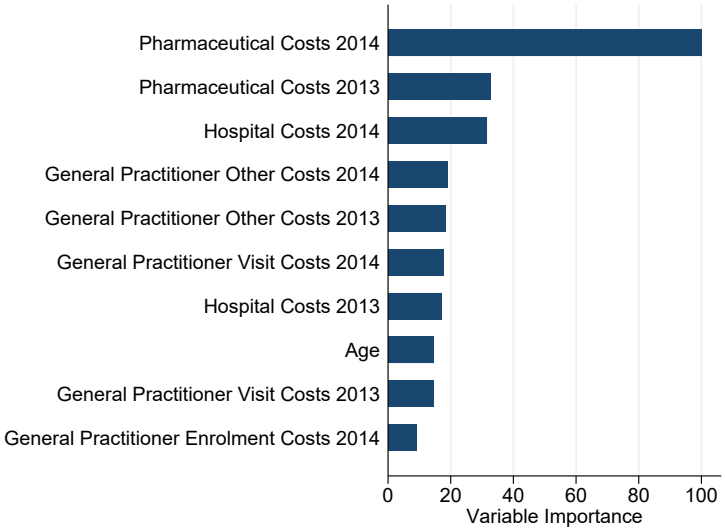
Appendix Figures

Figure 13: Calibration Plot of Prediction Model



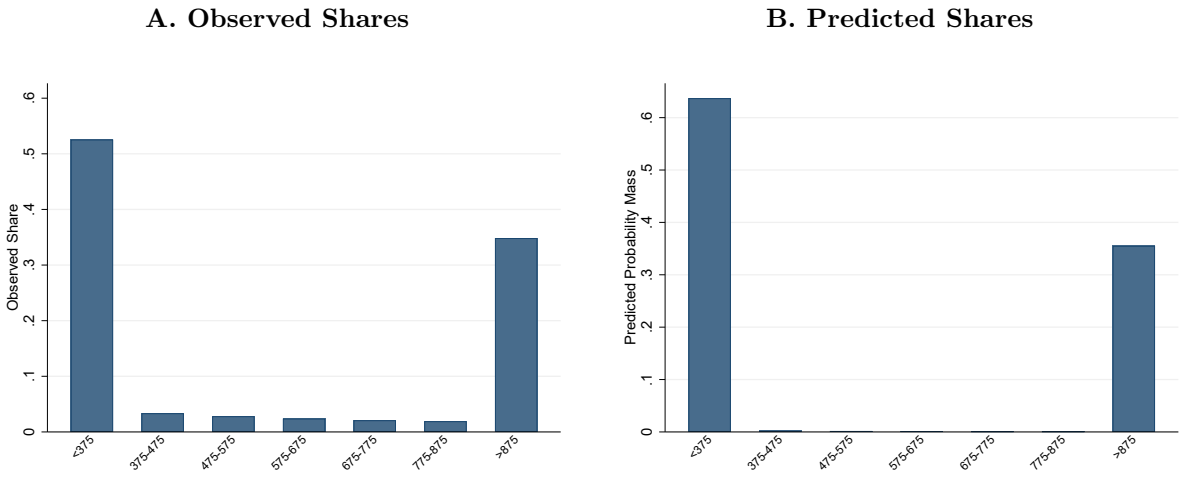
Notes: This figure is a binned scatterplot of the obtained predicted probabilities from the ensemble learner model (see Appendix A). This figure is made by plotting, for equal sized bins of predicted probabilities, the fraction of individuals with health costs below 375 EUR. As the binned probabilities match the 45-degree line, this figure shows that the model is well calibrated.

Figure 14: Variable Importance in Prediction with Random Forest



Notes: This figure shows the importance of selected variables in the prediction of health cost risk using only a random forest model. Variable importance is measured by the mean decrease in gini, ie. the average of a variable's total decrease in node impurity, weighted by the proportion of samples reaching that node in each individual decision tree in the random forest.

FIGURE 15: COST PREDICTIONS WITH MULTIPLE DEDUCTIBLE CATEGORIES



Notes: Panel A plots the observed share of individuals with health costs in all the deductible health cost brackets in 2015. Panel B plots the predicted shares of individuals in all deductible health cost brackets, where the prediction is from a random forest with the same predictors as described in Section IV.