Does the framing of patient cost-sharing incentives matter? The effects of deductibles vs. no-claim refunds*

Arthur P. Hayen, Tobias J. Klein, and Martin Salm†

April 2018

Abstract

In light of increasing health care expenditures, patient cost-sharing schemes have emerged as one of the main policy tools to reduce medical spending. We show that the effect of patient cost-sharing schemes on health care expenditures is not only determined by the economic incentives they provide, but also by the way these economic incentives are framed. Patients react to changes in economic incentives almost twice as strongly under a deductible policy than under a no-claims refund policy. Our preferred explanation is that individuals are loss-averse and respond differently to both schemes because they perceive deductible payments as a loss and no-claim refunds as a gain.

Key words: Patient cost-sharing, health insurance, framing, loss aversion.

JEL-classification: I13, D91, H51.

*First version: January 2017. We are grateful for the numerous insightful discussions we had with Johan Polder. We would also like to thank Viola Angelini, Arthur Attema, Jan Boone, Rudy Douven, Hans van Kippersluis, Jürgen Maurer, Pietro Tebaldi, Valentina Tonei, Albert Wong, conference participants at the RIVM/Celsus symposium in The Hague, the 2017 Netspar International Pension Workshop in Leiden, the workshop on Health Economics and Health Policy in Heidelberg, the LOLA Health Economics Study Group meeting in Rotterdam, and the IHEA congress in Boston, as well as seminar participants at Erasmus University Rotterdam, Tilburg University, the Hamburg Center for Health Economics, and the Technical University in Zurich for many helpful comments and suggestions. Funding of the PACOMED project by the RIVM is gratefully acknowledged.

†Hayen: Tranzo, Tilburg University; Klein: Netspar, CentER, Tilburg University; e-mail: T.J.Klein@uvt.nl; Salm: Netspar, CentER, Tilburg University.

1 Introduction

High and increasing health care expenditures are a major policy concern in many countries. One of the main policy tools to reduce medical spending is to introduce or expand patient cost-sharing schemes that let patients pay for a portion of health care expenditures not covered by health insurance. Across OECD countries, patients now pay on average around 20 percent of total health care expenditures out-of-pocket (OECD Indicators, 2015). Evidence from both field experiments (Manning et al., 1987; Finkelstein et al., 2012) and natural experiments (Chandra et al., 2010; Brot-Goldberg et al., 2017) shows that individuals reduce medical spending in response to economic incentives that follow from cost-sharing schemes such as deductibles and co-payments.

Yet, individuals might react not only to the economic incentives themselves, but also to the way these economic incentives are framed. For example, instead of framing cost-sharing incentives as a price that patients have to pay for health care utilization, they can also be framed as a refund that patients receive if they use little or no health care. From other contexts we know that individuals' choices can be affected by the framing of economic incentives (Schmitz and Ziebarth, 2016; Brown et al., 2016).[1]

More generally, seemingly small changes in policy schemes can have large effects, and understanding whether such changes make a difference is a key input in the design of economic policy (Duflo, 2017). In our context, the framing of cost-sharing incentives can have implications for cost containment, access to care, and financial risk born by patients. Understanding how the framing of cost-sharing incentives affects health care expenditures can help improve the design of health insurance plans.

In this study, we compare patients' responses to a deductible to the responses to a no-claim refund, two alternative patient cost-sharing schemes. The economic incentives under a deductible and a no-claim refund are very similar, but they are presented in a different way. Under a deductible policy, individuals pay out-of-pocket for all medical care up to the deductible limit. Under a no-claim refund policy, individuals receive a payment at the end of the year if their health care spending during the year was below the no-claim refund limit. The payment is equal to the difference between the no-claim refund limit and annual health care spending. Loss aversion implies that individuals respond stronger to losses than to gains (Kahneman and Tversky, 1979). If individuals perceive deductible payments as losses and lower no-claim refunds as foregone gains then we might expect that individuals will react stronger to deductibles than to no-claim refunds. Therefore, the different framing of cost-sharing incentives might have an impact on health care expenditures.

Usually, it is hard to study the effect of such differences in plan design, since patient populations differ between insurance plans. In our paper, we make use of the fact that in the Nether-

[1] Schmitz and Ziebarth (2016) show that framing of prices as euro amounts instead of a percentage of income has a large impact on the choice of health insurance plans in Germany. Brown et al. (2016) show that social security benefit claiming decisions are strongly affected by framing.

lands, both schemes have been in place at different points of time while the patient population and the services covered by health insurance remained comparable. In the years 2006 and 2007 Dutch law has mandated that health insurance contracts included a no-claim refund, and from the year 2008 onward, it required that health insurance contracts include an annual deductible.

Our analysis is based on administrative data from the Netherlands for the years 2006-2015. Our data are from a Dutch health insurer and include claim-level information on health care utilization for a general population of insured individuals, which we aggregate to around 9 million person-month observations. Having information on health care utilization at a monthly level is important for our empirical strategy, which exploits variation in cost-sharing incentives within a year.

Conceptually, one can think of cost-sharing as imposing a price of care. Under both a deductible policy and a no-claim refund the price for health care utilization can vary over the course of the year depending on whether or not an individual has exceeded her deductible or no-claim refund limit. For health care expenditures below these limits the price is one, while for expenditures above these limits the price is zero from the patient's perspective. Following Brot-Goldberg et al. (2017) we define prices as the price for the first unit of care in a given month, and we examine how the reaction to prices differs between the years when a no-claim refund policy was in place and the years when a deductible policy was in place. Our main hypothesis is that individuals respond stronger to prices under a deductible policy than under a no-claim refund policy.

We account for the possible endogeneity of prices with an instrumental variables approach. As instrumental variable for the price we use a simulated average price for people in the same risk score decile and age group, and with the same gender. This approach exploits differences in the evolution of prices across groups of individuals: individuals with higher risk scores tend to exceed cost-sharing limits more frequently and earlier in the year than individuals with lower risk scores. Our instrumental variable approach is similar to the approach by Ellis et al. (2017), who use average prices for different health insurance plans as instruments, leaving out the price of the respective individual. Instead of using average prices, we simulate prices using draws from the empirical distribution of health care costs within groups, pooling over all years and months. Thereby, we purge the simulated average prices of possibly confounding factors that may lead to a violation of the exclusion restriction. By design, variation in our instrument within the year is not affected by seasonal trends in health care expenditures and variation across years is solely driven by changes to the institutional rules, but not by the reaction of individuals to those changes. Comparable simulated instrumental variable approaches have been employed in other settings for example by Currie and Gruber (1996a,b) and Cutler and Gruber (1996), but are new in the context of studying the effects of patient cost-sharing.

We find that a higher price decreases health care expenditures under both a no-claim refund and a deductible policy, but the decrease is much larger under a deductible policy. Monthly health care expenditures are reduced by 17.5 percent under a no-claim refund and by 36.2

percent under a deductible. Hence, the framing of incentives can be quantitatively as important as the incentive itself. While these estimates might seem large, our estimates for the deductible scheme are similar to the results obtained by Brot-Goldberg et al. (2017). We do not find significant differences in the size of the effect by age and by income in the neighborhood, but we do find that women react stronger to prices than men and that effects are stronger for individuals with lower risk scores than for individuals with higher risk scores. Furthermore, cost-sharing has no economically significant effect for a range of high value treatments such as medication for diabetes and hypertension. This is not altered by the framing of cost-sharing incentives. Our results are robust to a number of sensitivity analyses that use alternative definitions of health care utilization and alternative specifications of explanatory variables. They cannot be explained by end of year effects, and in a placebo test we find no significant effect when we apply our empirical approach to 15-17 year olds who in the Netherlands are excluded from both no-claim refunds and deductibles.

We interpret our findings as evidence for a strong effect of the framing of cost-sharing incentives on health care expenditures. An alternative explanation would be the different timing of payments under a no-claim refund and a deductible policy. Invoices for deductible payments are sent a few months after treatment, while no-claim refunds are due only after the end of the calendar year. However, we show that the difference in responses to a deductible policy and a no-claim refund policy is too large to be explained by the difference in timing of payments. Based on a simple model similar to Einav et al. (2013) we show that the annual discount rate that would be necessary to explain our results is above 300%, which we consider to be implausibly large. We also show that if we discount payment streams with a large, but still plausible annual discount rate of 10% our results stay essentially unchanged.

Our study makes several contributions. First, our study contributes to the literature on the effect of patient cost-sharing on health care expenditures. Studies on the effect of patient cost-sharing on annual health care expenditures in the Netherlands include Ecorys (2011), Lambregts and van Vliet (2017), and Remmerswaal et al. (2017).[2] A focus of the recent literature on patient cost-sharing and health care expenditures is a better understanding of the behavioral underpinnings of patients' responses to cost-sharing schemes, e.g. to what degree patients respond to the dynamic incentives provided by cost-sharing schemes (Einav et al., 2015; Abaluck et al., 2015; Dalton et al., 2017), whether co-payments cause patients to shop for cheaper providers (Brot-Goldberg et al., 2017), or how responses vary by type of medical service (Ellis et al., 2017). In our study, we expand the knowledge about patients' responses to cost-sharing schemes in a new direction by showing that the framing of patient cost-sharing —as a deductible or as a no-claim refund—can have a large impact on health care utilization. In related research, Remmerswaal et al. (2017) show that the difference in annual health care expenditures in the Netherlands between 19 and 17 year olds is bigger in years in which a deductible was in place than in years when a no-claim refund was in place.

[2]A survey on the Dutch literature can be found in Remmerswaal et al. (2015).

Second, our study contributes to the literature on prospect theory and loss-aversion (Kahneman and Tversky, 1979). Already Johnson et al. (1993) predict that individuals should react stronger to deductibles than to no-claim refunds in insurance markets. However, van Winsen et al. (2016) point out that this prediction depends on a specific choice of reference point. If individuals choose their financial situation after paying insurance premiums as reference point then they will see a deductible payment as a loss and no-claim refund as a gain, and as a consequence they will react stronger to a deductible than to a no-claim refund. If, however, individuals choose their financial situation before paying insurance premiums as reference point, then they will feel to be in the loss domain under both schemes, and there should be no stronger response to deductibles than to no-claim refunds. Theory provides no clear guidance about how individuals choose reference points in the context of loss aversion. Reference points can be chosen as either the decision maker's *status quo*, an expectation, or an aspiration level (Baucells et al., 2011), and evidence on how reference points are chosen in field settings is scarce. The fact that we find a stronger response under a deductible than under a no-claim refund provides support for the hypothesis that individuals choose their financial situation after paying insurance premiums as reference point. Thus, our study contributes to our understanding of how individuals choose reference points in the context of patient cost-sharing.

Third, our study focuses on no-claim refunds, a form of cost-sharing in insurance markets that has received relatively little attention in the previous literature. No-claim refunds can be applied to many different types of insurance contracts with cost-sharing, and they are common for example in private health insurance plans in Germany. Previous studies provide some reasons why no-claim refunds might have advantages compared to more traditional cost-sharing schemes such as deductibles and co-payments. Johnson et al. (1993) argue that loss-averse agents might prefer an insurance plan with an no-claim refund over an insurance plan with a deductible. Individuals might see no-claim refunds as a "friendlier" form of cost-sharing. Furthermore, Fels (2017) shows in a theoretical model that incentivizing efficient health care utilization with a bonus rather than with deductibles and co-payments can be optimal if consumers make mistakes and can be financially constrained.

In the following, Section 2 describes the institutional background. A conceptual framework based on loss aversion is proposed in Section 3. Section 4 provides details on the data. Section 5 presents descriptive evidence. Section 6 describes the empirical approach. Section 7 presents the main results, and Section 8 presents robustness checks. Section 9 discusses policy implications and concludes.

2 Institutional background

The Dutch health care system lends itself well for separating framing effects from other determinants of demand. Insurance plans with a no-claim refund in the years 2006 and 2007 and insurance plans with a deductible in the years thereafter were markedly similar in terms of coverage and services for which there was cost-sharing. These plans also provided similar economic incentives because the deductible limit was similar in size to the no-claim refund limit. At the same time, there were no substantial changes on the supply side.

Next, we provide some background information on the health care system in the Netherlands, the two cost-sharing schemes, and the way in which providers are paid.[3]

2.1 Health insurance in the Netherlands

The Netherlands has a system of universal, mandatory health insurance coverage. Since a reform of the health insurance system in the year 2006, insurance is funded in about equal proportions by income-dependent employer contributions and premiums paid by the insured. At the same time, there is a risk equalization scheme between insurance providers.

Every resident of the Netherlands is obliged to purchase a basic, nationally standardized insurance plan on the market, from one of several competing private health insurers. The premiums are community-rated, meaning that the insurance premium is similar across patients, irrespective of characteristics like age, gender, and health status. Premiums may only differ for different levels of cost-sharing, and for enrollees who enroll via a collective bargaining agreement (in which case they can receive a discount of up to a 10% of their monthly insurance premium). Furthermore, health insurers are not allowed to deny basic health insurance coverage.

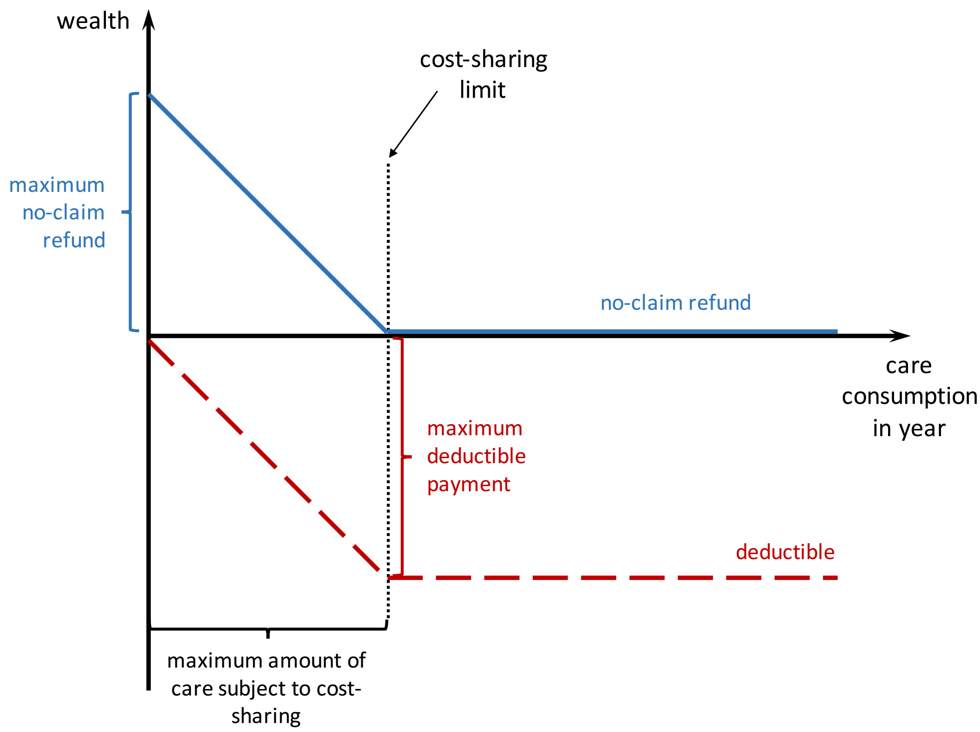
The basic health insurance plan includes a wide range of services, such as general practitioner (GP) services, hospital care, prescription drugs, mental health care, and medical devices such as hearing aids and prostheses.[4] On top of this basic insurance package, individuals can buy additional coverage on the market for supplemental health insurance, for example for dental care.

Private insurers compete with one another in terms of policy conditions. Apart from setting the level of the monthly premium, health insurers may also restrict coverage, for example by contracting selectively with providers. However, case law demands that health insurers still pay a "substantial" amount of the bill in case a patients visits an out-of-network provider. The contracts between health insurer and care providers typically contain agreements on prices, volumes, and quality (Krabbe-Alkemade et al., 2016). At the start of each year, insurees are allowed to switch to another health insurer. Between 2006 and 2012, the annual switching rates ranged between 4 and 6 percent (NZA, 2012).

[3] See for instance Schäfer et al. (2010) for additional details.

[4] Services included in the basic health insurance plan changed slightly over the years. The most notable change was the inclusion of mental health care services in the year 2008. As a robustness check we estimate a specificiation in which we exclude mental health care services.

Figure 1: No-claim refunds and deductibles



Notes: This figure shows financial implications of care consumption under a no-claim refund and a deductible. The cost-sharing limit is the same for both.

2.2 Patient cost-sharing

Since 2006 the Dutch government has implemented two alternative patient cost-sharing schemes. They have been in place at different points in time. Both schemes were mandatory and uniform across health insurance plans. In the years 2006 and 2007, health insurance plans had to include a no-claim refund, and since the year 2008 health insurance plans have to include an annual deductible. Under the no-claim refund scheme enrollees receive a refund at the end of the year if their annual health care spending during the year was below the no-claim refund limit. The amount of the refund is equal to the difference between the no-claim refund limit and annual health care spending. Under the deductible scheme individuals pay out-of-pocket for all care up to the deductible limit. Health care utilization beyond the deductible limit is free from the patient's perspective.

Annual payment streams for both schemes are illustrated in Figure 1. For illustration purposes, the no-claim refund limit in the figure is set equal to the deductible limit. The horizontal axis depicts annual health care spending. The vertical axis shows annual payments relative to a reference point that in the figure is set to the financial situation after paying health insurance premiums. Under a no-claim refund scheme wealth is always above the reference point for any level of annual health care spending. Under a deductible scheme wealth is always below the reference point. The two curves for no-claim refunds and for deductibles have the same shape, but they are shifted vertically. The identical shape of the two curves reflects that the "price" that patients have to pay for an additional euro of health care utilization is identical under both schemes for any level of annual health care spending. The vertical shift between the two curves does not imply that health care is cheaper under a no-claim refund scheme than under a deductible scheme. Instead, the vertical shift will be offset by different levels of health insurance premiums, which will be higher under the no-claim refund scheme than under the deductible scheme.

The no-claim refund limit in the years 2006 and 2007 was 255 euros. The deductible limit started at 150 euros in the year 2008, and it was subsequently increased in several steps to 375 euros in the year 2015.[5] Minors up to the age of 18 years are excluded from patient cost-sharing under both the no-claim refund and the deductible scheme (a feature we make use of in Section 8.1 to conduct a placebo test). Furthermore, certain types of medical services such as GP care and maternity care are also excluded from patient cost-sharing (GPs are a patient's first point of contact when in need for health care, and they have a gatekeeping function in the Dutch health care system, cf. Schäfer et al., 2010).

Under both regimes, individuals could opt for a voluntary, additional deductible (in quantities of 100 euros, with a maximum of 500 euros).[6] In return, they received a discount on their monthly premium. Under the no-claim refund, the no-claim limit needed to be exhausted first, before health care spending would count towards the voluntary deductible.[7]

Even though cost-sharing rules apply uniformly to all patients, people with a chronic illness can apply for a subsidy that partially compensates them for their higher expected deductible payments. Eligibility is determined based on whether or not people have received treatment for chronic illnesses in the previous two years. Since it is based on expenditures in previous years, the subsidy does not depend on current year's spending. Despite the fact that many people qualify for this subsidy, a substantial share of the eligible population is not aware of the rule (Reitsma-van Rooijen and Jong, 2009).

2.3 Payment of providers and timing of deductible payments

The way in which providers are paid is important for understanding the spending dynamics under cost-sharing regimes, because it influences the timing of payments by the patients. In the Netherlands, providers are predominantly paid for bundles of services (Schäfer et al., 2010). For specialist care, services are bundled into so-called "diagnosis-treatment combinations (DTC's)" (Hasaart, 2011; Van de Ven and Schut, 2009). The specialist assigns the patient to a DTC upon

[5] The deductible limits were 155 euros in 2009, 165 euros in 2010, 170 euros in 2011, 220 euros in 2012, 350 euros in 2013, and 360 euros in 2014.

[6] The number of individuals doing so was almost negligible. We exclude them from our analysis. See Section 4 for details.

[7] GP care is generally excluded from cost-sharing, but did fall under the voluntary deductible when there was a no-claim refund.

first contact, but is allowed to adjust this choice later should there be changes in the treatment process (cf. Hasaart, 2011). One consequence of bundled payment is that patients are not billed for each service separately, but for the entire DTC at once.

Deductible payments are not collected by providers at the point of service, but by the health insurer. This means that a health service needs to be billed first to the health insurer before the deductible payment is due. This is usually done with a delay (the average delay across all claims in our data is 60 days, see also Figure C.5 in the Online Appendix). Thereafter, when the health insurer receives the bill, she will determine the amount the patient has to pay based on the remaining deductible. In our analysis below, the key explanatory variable is whether, because of other treatments, cost-sharing limits have been exhausted at the time a new treatment decision is made. Then, we say that the price of care is zero for the individual.

Importantly, following Brot-Goldberg et al. (2017) and Ellis et al. (2017), we count the expenditures on other treatments towards the cost-sharing limit if these other treatments have been started at the time a new treatment decision is made, independent of whether or not the other treatments have already been billed to the health insurer. This implies that patients might not always know whether or not they have exceeded their cost-sharing limit at a given point in time, because they do not know the exact costs for treatments that have been started, but have not yet been billed to the health insurer. However, for more frequent treatments patients might know the approximate costs based on past experience. For serious treatments, which for example involve repeated consultations with a medical specialist or an overnight stay at an hospital, patients might not know the exact costs, but they might be aware that the costs of such treatment typically exceed the relatively low cost-sharing limits in the Netherlands.

3 Conceptual framework

3.1 Non-linear price schedules

Our study examines how patients react to cost-sharing under a deductible and no-claim refund policy, respectively. To study this, we use the concept of the price of health care, which is one if an individual has to pay herself and zero if a treatment is fully covered by the insurance.

Under both policies individuals make decisions throughout the year. The price of care is a function of past health care consumption in the same year, and it can thus vary over the course of a year. At the beginning of the year the price of care is one. After patients have exceeded their annual cost-sharing limit the price they pay for additional care is zero. Forward-looking decision-makers should take this into account when making choices about health care utilization. In a seminal theory article, Keeler and Rolph (1988) present a set of assumptions including risk neutrality and separability of the utility function in health and money under which the expected end-of-year price—this is equal to one minus the probability of hitting the cost-sharing limit by the end the year—is the only relevant price. Ellis (1986) refines this by also incorporating risk aversion and derives the relevant shadow price. Based on this, one could proceed and spell out a structural model and estimate it using our data. This will involve making a set of assumptions on the stochastic process describing the evolution and temporal dependence of health care needs. We do so in a companion paper, Hayen et al. (2017), in which we study the reaction of patients to the deductible and then perform counterfactual experiments.

Here, we instead follow Brot-Goldberg et al. (2017) and Ellis et al. (2017) and focus on the question of how health care consumption depends on the current price of health care at the point in time at which the decision is made, and we then study how this reaction depends on the framing of cost-sharing. We do so for two related reasons. First, recent empirical studies find evidence for substantial myopia in patients' decisions about health care use in the presence of non-linear price schedules. In particular, Brot-Goldberg et al. (2017) show that patients react to current prices even after controlling for expected end-of-year prices.[8] Second, health care consumption is lumpy and deductibles are relatively small in the Netherlands. Both the current price and the expected end-of-year ("future") price are functions of past health care consumption, and because of the lumpiness and the small cost-sharing limits both are highly correlated. In our estimation sample, the correlation is 0.8.

In addition, we intentionally do not control for the *evolution* of the future price over the course of the year, as it is not the focus of this paper whether the reaction we observe is a reaction to the future price or the current price.[9] We will however control for differences across individuals by controlling for the future price at the beginning of the year. Further details are provided in Section 6.

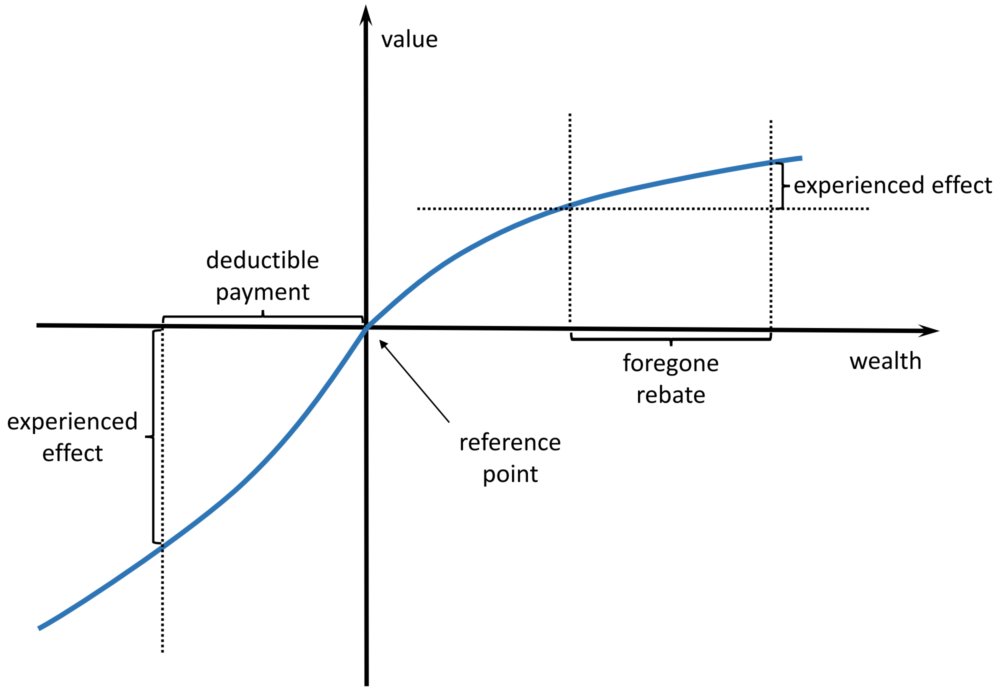
3.2 Framing as gains and losses

Loss aversion, one of the elements of prospect theory, postulates that decision-makers value losses more than corresponding gains (Kahneman and Tversky, 1979). This concept can also be applied to patient cost-sharing for health insurance where cost-sharing incentives can be perceived either as a loss or a gain. Whether patients perceive payments as a loss or a gain depends on their reference point. Previous theoretical studies propose alternative reference points in the

[8]Also Abaluck et al. (2015) and Einav et al. (2015) find evidence for myopia for Medicare Part D patients. Abaluck et al. (2015) find that a model with substantial myopia fits the data well. Einav et al. (2015) estimate a discount rate of around 7 percent per week, which means that the yearly discount rate is 98 percent. Their preferred interpretation of this model parameter is that it is "a behavioral parameter that also reflects individuals' understanding of the insurance coverage contract, in particular the salience of the (future) nonlinearities of the contract."

[9]In fact, the coefficient on the current price will capture the effect of both, the evolution of the current price *and* the evolution of the future price. Qualitatively, we will see a similar reaction in either case, as prices are highly positively correlated. Because of the correlation, our estimates will put the more weight on the current price the more individuals discount. To see this, consider the two polar cases. Suppose individuals are fully myopic. Then, they will react only to the current price, which is included in the regression. Instead assume that they are fully forward-looking and don't discount. Then, they will react only to the future price. As this price is highly positively correlated with the current price that is included in the regression, we will qualitatively capture the effect of the future price even if we include the current price into the regressions.

Figure 2: Experienced effects



Notes: This figure shows the experienced effects of a foregone refund and of a deductible payment. The amount for both is the same.

context of cost-sharing schemes in insurance markets. Johnson et al. (1993) consider the case when patients perceive their financial situation after paying insurance premiums as reference point. Compared to this reference point they will view a deductible payment as a loss. On the other hand, if they receive a refund at the end of the year under a no-claim refund policy, then they will view this payment as a gain relative to their reference point. Thus, patients should react stronger to cost-sharing under a deductible policy than under a no-claim refund policy. The exact same annual payment can be evaluated differently when cost-sharing incentives are provided by a deductible instead of a no-claim refund.

The reasoning above is illustrated in Figure 2, which shows a utility curve ("value") under loss aversion when the cost-sharing limit has not been reached. The reference point is the financial situation after paying health insurance premiums. The utility curve is steeper to the left than to the right of the cutoff point, and in both directions the utility curve becomes flatter with increasing distance to the reference point. The figure shows that the experienced effect of a deductible payment is higher than the effect of an equivalent reduction in the refund under a no-claim refund policy. When patients make decisions about medical care use they will compare the value of the treatment with the value they assign to their own payment, either a deductible payment or a foregone refund under a no-claim refund policy. Thus, patients will use less medical care under a deductible policy than under a no-claim refund policy.[10]

To be precise, the hypothesis stated above is that individuals are loss-averse *and* that patients use their financial situation after paying insurance premiums as reference point. Individuals might also choose reference points other than their financial situation after paying insurance premiums. van Winsen et al. (2016) point out that the conclusion above does not hold if patients use the financial situation before paying health insurance premiums as reference point. In this case, patients would always be in the loss domain under both a deductible policy and a no-claim refund policy and for any level of annual health care expenditures, and we would not expect a stronger response to deductibles than to no-claim refunds.

Previous studies provide no guideline which reference point patients choose in the context of cost-sharing schemes in insurance markets. Finding that patients react stronger to deductibles suggests that at least some patients are loss-averse and that their reference point is given by wealth *after* insurance premiums have been paid.

4 Data

Our data come from a large Dutch health insurer. They were obtained under a concurrent pilot project in which a new payment model for GPs was designed and implemented (see Hayen et al., 2015, for details). This pilot ran from July 2014 to the end of 2015, but the data also include pre-intervention years. The data cover the years 2006 to 2015.

From a set of approximately 200 GPs that participated either in the intervention or were part of the control group, we first selected all insurance enrollees who were registered with those GPs. For our main analysis, we used observations for individuals who were at least 19 years old (such that they were subject to cost-sharing for the entire year—information on the month of birth is not available to us), were insured for the entire year (which means that they did not die before the end of the year, as one is allowed to change insurance only at the start of the year), and did not have a voluntary deductible. This gives us an unbalanced panel with about 85,000 insured individuals in any given year.[11]

To assess the representativeness of our sample, we compare the average age and the proportion of women, both among individuals who are at least 19 years old, to the respective values for the population. Differences are small. Individuals in our sample are about 1 year older on average and the percentage of women is by one percentage point higher. Moreover, exclud-

[10]In our empirical analysis, for reasons we have given in Section 3.1, we study the reaction to current prices. The analysis is carried out on a monthly level. The reasoning in the paragraph above also applies to the monthly level: Figure 2 then shows utility defined over gains and losses at the monthly level, when the reference point is wealth after the monthly premium has been paid.

[11]People who met this requirement were both insured with the health insurer, and registered with one of the GPs in our sample. The panel is unbalanced because individuals may have changed insurer or may have chosen a voluntary deductible in some years (and were therefore excluded in those years). In our analysis below, we control for time effects and thereby also for differences in sample composition over the years. Individuals in our sample were always observed for entire years and our main parameters of interest are estimated from within-year variation.

ing individuals with a voluntary deductible from our sample does not seem to cause serious concerns, as we exclude only about 2 percent of observations for that reason.[12]

Our data include information on each person's insurance policy, length of enrollment, and medical care utilization at the claims level. These claims-level data are very detailed. Among others, they include the billing code of the particular service, a full description of the service, the full DTC for medical specialist care, the date on which the care was delivered, the billing date (by the provider to the insurance company, not the billing date to the patient), the amount paid by the insurance and the amount paid out-of-pocket by the patient. For prescription drugs, we know the exact quantity delivered (e.g. 5 ml of Tobramycin) and their classification according to the Anatomical Therapeutic Chemical Classification System with defined daily doses (World Health Organization, 2006).

We inferred information on whether or not an individual has one of 25 chronic conditions (e.g. type 2 diabetes) based on claims for prescription drugs. In case a person has billed more than 180 daily doses of medicines linked to the treatment of a particular chronic condition, she was believed to have this condition.[13]

The data also contain demographic information (age, gender), and a person's postal code at the 6-digit level.[14] We use the latter to add information on socioeconomic status: we add income (as measured in December 2008) and the percentage of non-western immigrants (as measured over 2010) at the 6-digit post code level using data from the Dutch Central Bureau of Statistics (Statistics Netherlands, 2012). We also add 2006 socioeconomic status scores at the 4-digit postal code level, as calculated by the Netherlands Institute for Social Research (Netherlands Institute for Social Research, 2006).[15] These status scores are calculated based on income, education, and employment status.

Outcome variables in our study are measures of health care expenditures. Our measures of health care expenditures include all care that is included in the basic health insurance package and is in principle subject to patient cost-sharing. The latter restriction excludes GP care and maternity care. Our two outcome variables in the main analysis measure monthly health care expenditures at the intensive margin and at the extensive margin, respectively. Our measure of health care expenditures at the intensive margin is given by $y_{it} = \log(spending_{it} + 1)$ where

[12] The fraction of individuals with a voluntary deductible varies over the years. In the beginning, it was about 1 percent and then grew to 4 percent over the course of 10 years.

[13] We obtained additional data from the Royal Dutch Pharmacists Association to translate each prescription drug claim into the number of defined daily doses obtained. For each individual, we then summed the number of defined daily doses obtained per medicine group as defined by their 6-digit Anatomical Therapeutic Chemical Classification (e.g. metformine) and used data from the Dutch Health Care Institute to link a medicine's classification to the treatment of chronic conditions (e.g. metformine is used for the treatment of type 2 diabetes).

[14] A 6-digit post code consists of a number with 4 digits and two letters. In 2007, there were 456,913 different post code areas (http://postcodeinfo.nl/files/factsheet-postcodeinfonl.pdf, accessed November 2017) and 17 million inhabitants (http://statline.cbs.nl/StatWeb/publication/?PA=37296ned&D1=a&D2=0,10,20,30,40,50,60,(l-1),l&HDR=G1&STB=T, accessed November 2017). This means that there are on average 37 inhabitants per post code area.

[15] There are 4054 4-digit post codes (https://home.kpn.nl/pagklein/postcodes.html, accessed November 2017). This means that there are on average 4193 inhabitants per post code.

$spending_{it}$ are the health care expenditures of individual i in month t.[16] Our measure of health care expenditures at the extensive margin is a binary indicator that equals 1 in case a person consumed a positive amount of medical care in a given month, and it equals 0 in case she did not comsume medical care in this month.

Our main independent variable in both models is the price p_{it} for person i in month t, which we define as the marginal cost-sharing rate for the first unit of care consumed within a month. This price equals 0 in case a person has exhausted her deductible or exceeded the no-claim refund limit at the beginning of a month, and 1 in case she has not. Whether or not medical spending is above the cost-sharing limit is determined based on the treatments that have been started up to that point of the year, independent of whether this care has already been billed to the health insurer (see discussion in section 2.3).

Additional control variables include age, gender, risk score deciles, and the expected end of year price at the beginning of January. The risk score is given by a person's predicted annual expenditures divided by the sample's average annual expenditures. We obtained this prediction by performing a linear regression of annual health care expenditures on several risk characteristics, which are age and gender (gender fully interacted with a third-order polynomial in age), last year's spending decile, indicators for chronic illnesses in the previous year, and quartile splits for income, the percentage of non-western immigrants, and the socioeconomic status score in the neighborhood. With this we aim to mimic the way risk scores are calculated in Dutch risk adjustment scheme between health insurers.[17] Based on individual risk scores we compute a set of risk score decile indicators at the person-year level. We also estimate the expected end-of-year ("future") price from the perspective of the beginning of the year (see also Section 3.1). This price is calculated as the average end-of-year price at the level of a cell formed of the risk score decile and age/gender category (18-40, 40-65, and over 65 years of age).

Table 1 presents summary statistics for three different samples. The first column is for our estimation sample. The second column is for the full sample of individuals in our data who are at least 19 years old. The third column is for the sample of individuals who are in our data in all years. About half of the individuals in our estimation sample are female, and they were in their late 40s on average. Average gross income at the postcode level is about 2200 euros per month. In a given month, average health care expenditures that are subject to cost-sharing are 168 euros, and the likelihood that there are positive expenditures in a given month is 42 percent.[18]

[16]See Online Appendix A.1 for details on how this was constructed.

[17]See Online Appendix A.2 for details. In the national risk-equalization scheme, insurance companies compensate one another for differences in their risk pools. This is necessary because risk-rating is not allowed. The actual risk scores that are used for this use additional information that was not available to us. For instance, they use income at the individual level, while we use income at the 6-digit post code level. We expect the difference to be small, as a post code covers only 37 persons on average (footnote 14).

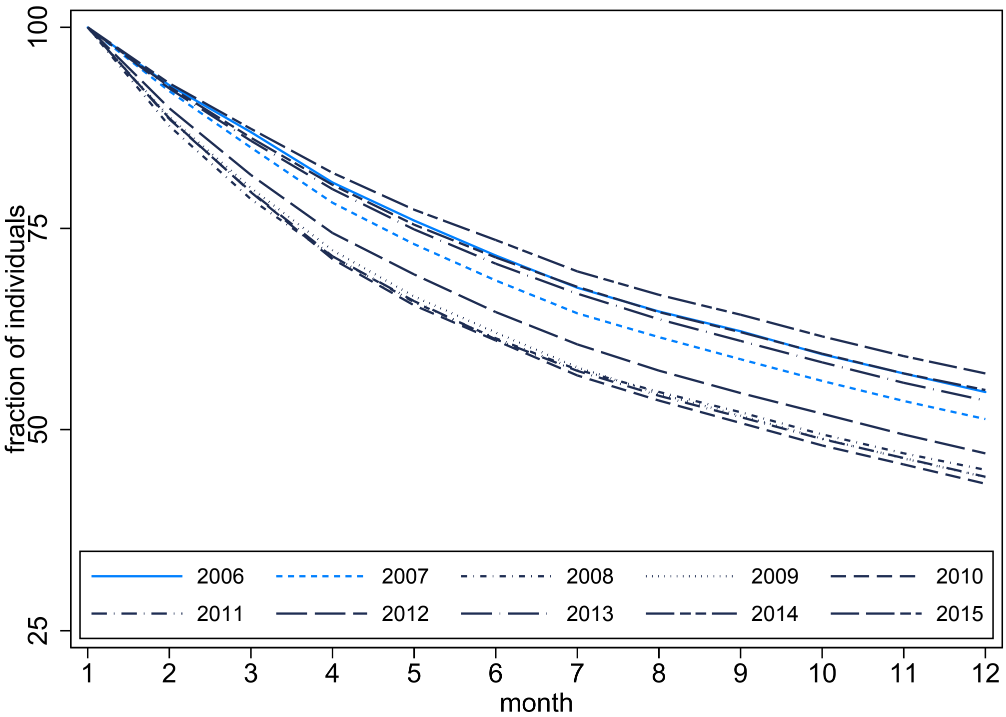
[18]Figure C.2 in the Online Appendix plots the consumption of care within the year. It shows that there is no strong seasonal component. Table C.2 presents average care consumption and the fraction who hit the cost-sharing limit by year, respectively.

Table 1: Summary statistics for different samples

	estimation sample	full sample	balanced sample
individual and neighborhood characteristics			
female	0.52	0.52	0.52
age	49.80	48.75	53.77
income (if not missing)	2182.36	2188.58	2167.33
indicator income missing	0.04	0.04	0.03
care consumption per month			
expenditure	168.22	161.94	162.22
has claim	0.42	0.41	0.44
number observations	8,772,000	10,965,504	3,855,276

Notes: This table shows summary statistics for individuals who are at least 19 years old, insured for the entire year, and choose the minimum deductible. One observation is a person-month. The first column is for all individuals in our estimation sample who automatically meet these requirements. The second column is for the full sample of individuals meeting these criteria. There are more observations because in our estimation sample we also require that information for the respective previous year is available. This information is used to calculate the risk score. The third column is for the balanced sample of individuals who have been at least 19 years old in 2007 and were in our data until 2015. All presented values are means across individuals and time. Statistics for the estimation and the balanced sample are reported for the years 2007 to 2015. Income is gross income at the 6-digit post code level in nominal euros. Care consumption is the consumption of care that falls in principle under the no-claim refund policy or the deductible. Has claim is one if care consumption is non-zero in a given month.

Figure 3: Fraction of individuals below deductible or no-claim refund limit

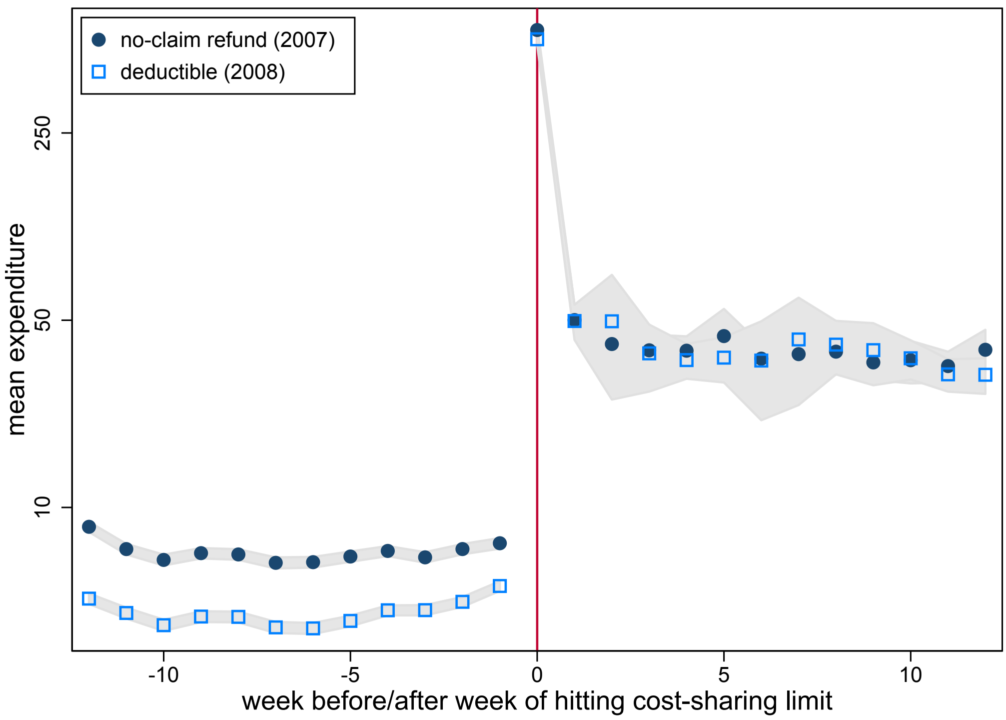


Notes: This figure shows the fraction of individuals who have not hit the deductible or the no-claim refund limit, by month and for all years between 2006 and 2015. Calculated for the full unbalanced panel. The cost-sharing limits were 255, 255, 150, 155, 165, 170, 220, 350, 360, and 375 euros.

Comparing means across columns shows that individuals in the balanced sample, who remained with both the insurer and the sampled GPs throughout our study period, are somewhat older on average. They have a higher probability to have a claim in a given month, but lower average health care expenditures. Still, differences across samples are not dramatic, especially given the fact that the balanced sample is less than half the size of the estimation sample.

Figure 3 shows how the average price of care evolves over the course of the year for the years in our study period. The horizontal axis shows months of the year. The vertical axis shows the fraction of individuals whose care consumption was still below the cost-sharing limits at the beginning of the month. These are the individuals for whom the price of care is one. By construction, the average price of care is monotonically decreasing over the course of the year. Level differences between curves are driven by a combination of differences in medical spending across years and differences in no-claim refund limits and deductible limits across years. One can see that by the end of each year about half of the individuals have exhausted the respective cost sharing limit.

Figure 4: Care consumption around week in which cost-sharing ends



Notes: This figure shows average health care expenditures in weeks before and after exceeding the no-claim refund limit or the deductible in 2007 and 2008, respectively. Cost-sharing incentives were framed as a no-claim refund in 2007 and as a deductible in 2008.

5 Descriptive evidence

The central question in this paper is whether individuals react differently to similar incentives when they are framed in terms of a no-claim refund instead of a deductible. Figure 4 presents our main model-free evidence documenting that they react stronger to a deductible than to a no-claim refund. For this, we follow individuals who exceed the no-claim refund or deductible limit in a given year, and we compare their health care spending in the weeks before and after hitting the cost-sharing limit. We do so for a balanced panel of individuals whom we observe at least 12 weeks before hitting the cost-sharing limit and 12 weeks thereafter in the same calendar year. This means that we exclude all observations for individuals who do not hit the cost-sharing limit during the year or who hit the cost-sharing limit very late or very early in the year.

The horizontal axis shows weeks relative to the week of hitting the deductible limit (or the no-claim refund limit). The week of hitting the cost-sharing limit is set to zero. The vertical axis shows average weekly health care expenditures. Each dot shows the average health care expenditures for a different week alternatively in the year 2007 when a no-claim refund was in place and in the year 2008 when a deductible policy was in place. To the right of the vertical line, after hitting the cost-sharing limit, the price of care is zero. We find that then, average expenditures are very similar under the no-claim refund in 2007 and the deductible in 2008. By construction, there is a peak in health care expenditures in the week that the deductible limit (or the no-claim refund limit) is hit, but in the weeks thereafter health care expenditures are essentially constant over a 12 week period. To the left, before hitting the cost-sharing limit, the current price is one, and individuals are subject to cost-sharing incentives. Under both schemes, average expenditures are lower to the left than to the right of the red line. However, expenditures are substantially lower in 2008 under the deductible than in 2007 under the no-claim refund. Accordingly, the increase in health care expenditures at the time of exceeding the cost-sharing limit is much larger under the deductible scheme than under the no-claim refund scheme.

Thereby, this figure shows strong suggestive evidence for our main hypothesis that health care expenditures respond stronger to deductibles than to no-claim refunds. We call this evidence suggestive because the increase in health care expenditures upon hitting the deductible cannot be interpreted as causal effect. It could also be explained for example by serial correlation in health care expenditures if high health care expenditures in the week of hitting the cost-sharing limit tend to be followed by other weeks with high health care expenditures. However, the figure shows that health care expenditures tend to level out soon after hitting the cost-sharing limit. Furthermore, serial correlation in health care expenditures can not explain why the responses to hitting the cost-sharing limit is stronger under the deductible than under the no-claim refund. In the following sections we turn to a more formal analysis.

6 Empirical approach

Our goal is to estimate the effect of the price of care on health care utilization, alternatively under a no-claim refund policy and a deductible policy. We specify the estimation equation

$$y_{it} = p_{it} d_t^{\text{no-claim refund}} \beta^{\text{no-claim refund}} + p_{it} d_t^{\text{deductible}} \beta^{\text{deductible}} + x_{it}' \gamma + \varepsilon_{it}, \tag{1}$$

where p_{it} is the price of care for individual i at the beginning of month t, $d_t^{\text{no-claim refund}}$ is a binary indicator for months in which the no-claim refund was in place, $d_t^{\text{deductible}}$ is a binary indicator for months in which the deductible was in place, and x_{it} is a vector of controls that includes a constant term. ε_{it} is an error term, and γ is a vector of parameters. $\beta^{\text{no-claim refund}}$ and $\beta^{\text{deductible}}$ are the main parameters of interest. They measure how health care expenditures respond to prices under a no-claim refund and a deductible, respectively.

In our baseline specifications, the dependent variable y_{it} is either the log of health care costs of individual i in month t plus 1 (as health care costs can be zero), so that coefficients are approximately percentage changes in total spending.[19] Or the dependent variable is a binary indicator for any expenditures in a given month, in which case coefficients measure effects at the extensive margin.

Determinants of health care costs in x_{it} include a full set of month and year dummies in order to control for seasonal variation in health care needs related to, for instance, the flu and for changes across years that are, for instance, driven by health care inflation. x_{it} also includes indicator variables for 5 year age bins and for gender. In addition, x_{it} includes indicators for the deciles of the individual risk score and the expected end-of-year price from the perspective of the beginning of the year (see Section 3.1). The risk score controls for differences in expected health care expenditures across individuals, and the expected end-of-the-year price captures in addition that cost-sharing limits and health care prices change over the course of our study period such that in different years individuals with the same risk score face a different probability of exceeding the cost-sharing limit and experiencing a zero price. Finally, x_{it} includes the log of lagged health care expenditures in the previous three months plus one in order to control for autocorrelation in health care expenditures.

We estimate the parameters using an instrumental variables approach, which we describe below. Throughout, reported standard errors are heteroskedasticity-robust and clustered at the level of groups defined by the risk score decile, three age categories, gender, and year. In our definition of clusters we follow the level of variation in our instrumental variable.[20]

Next, in Section 6.1, we discuss potential threats to identification. Then, in Section 6.2, we explain how we address these concerns using our instrumental variables approach.

[19] See Section 7 for further discussion.

[20] We have merged the two higher age categories for the lowest 4 risk score deciles. Therefore, there are 468 clusters (10 deciles times 3 age categories times 2 genders minus 8 merged groups) in our baseline sample.

6.1 Threats to identification

p_{it} is individual-month specific, and it is a function of past care consumption by the same individual within the same calendar year. p_{it} is zero when individuals have exceeded the cost-sharing limit and one otherwise. This means that a first concern could be that our estimates may suffer from omitted variables bias when we do not sufficiently control for individual differences in expected spending.

To see this, suppose that the omitted variable was the risk score (for which we actually control in our analysis, among others). For a given p_{it}, individuals with a high risk score will on average consume more care compared to those with a low risk score. Because prices are a function of past health care consumption, p_{it} will therefore on average be lower for individuals with a higher risk score, which means that a regression of health care consumption on prices will lead to an upward-biased estimate of $\beta^{\text{no-claim refund}}$ and $\beta^{\text{deductible}}$ when we do not control for the risk score. This concern is partly already addressed by including numerous control variables for health care needs in x_{it}, including the risk score. The remaining concerns are addressed by performing an instrumental variables approach that is detailed below.

A second concern is also related to the fact that p_{it} is a function of past health care consumption. The price in t depends on past values of ε_{is}, when $s < t$. If there is autocorrelation in the error term, for example because health shocks are correlated over time, then p_{it} will also be correlated with ε_{it}, because p_{it} is a function of ε_{is}, while ε_{it} is correlated with past ε_{is}. This will result in prices being endogenous.[21] We expect this to be less of a concern in the Netherlands than in other countries, because services are not charged separately, but bundled into DTC's, as described in Section 2.3. Still, following Brot-Goldberg et al. (2017), we control for health care expenditures in the previous three months and address remaining concerns by performing instrumental variables estimation.

A third potential concern is that seasonal trends in health care consumption differ between groups of the population. In equation (1) we control for period effects, but not for interactions of period effects with individual characteristics. It could for instance be that younger individuals are more likely than older individuals to have skiing accidents, which tend to occur in the winter months at the beginning of the year. A consequence of this would be that prices are negatively correlated with the error term, as the error term also captures the deviation of the group-level seasonal trend for younger individuals from the general seasonal trend, which is captured by period fixed effects. Also this concern is addressed with our instrumental variables approach.

[21] Note that this means that we cannot use a fixed effects estimator to control for level differences, because a fixed effects estimator would only yield consistent estimates if strict exogeneity would hold. This does not hold here because the regressor p_{it} is correlated with some ε_{is}, $s < t$, for the same individual.

6.2 Instrumental variables approach

The mechanical nature in which prices depend on past health care consumption can be used to construct an instrument. Price changes over the course of the year are related to health care expenditures. Thus, groups in the population with different levels of expected health care expenditures tend to face different prices. For example, for individuals with high risk scores prices tend to decline faster over the course of the year than for individuals with low risk scores. In our empirical approach we use the simulated average price at the group level as an instrumental variable for the individual price p_{it}, and we exploit changes within the year in the fraction of individuals who are subject to cost-sharing as a source of exogenous variation.

This can be seen as a generalization of a differences-in-differences estimator. For illustration, suppose that there are only two groups and two points in time and that for both groups the average price changes between the two points in time, but by a different amount. In addition to the change in prices, for both groups we can also compute the change in average spending between the two points in time. Then the instrumental variable estimator relates the difference in the change in spending between the two groups to the difference in the change in prices.

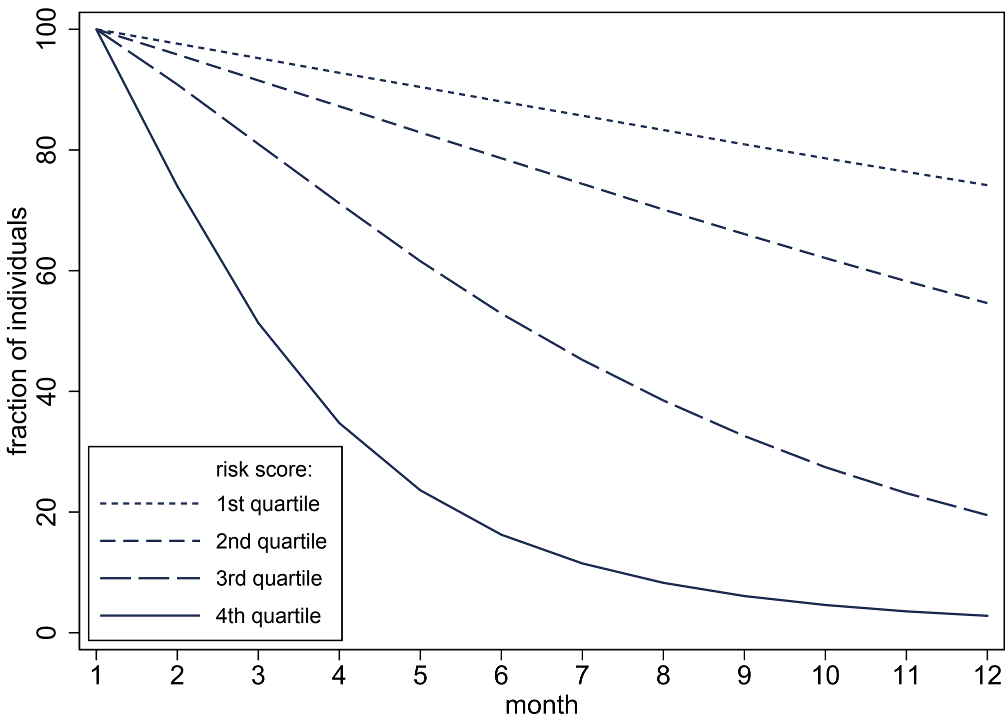
In a similar context, Ellis et al. (2017) use the average price at the group level as an instrument for the price, leaving out the observations for individual i, respectively, and controlling for differences in the level of spending across groups. Our approach is similar, but differs in two main aspects. First, Ellis et al. (2017) specify a group to consist of all individuals who chose a particular insurance plan. We instead define a group by the combination of the risk score decile, an age group, and gender.[22] Second, instead of the fraction of individuals who have not hit the cost-sharing limit at time t, we use the *simulated* fraction of individuals who have not hit the cost-sharing limit at time t.

We simulate this fraction by drawing from the empirical distribution of monthly health care consumption across all months and years, but for the same group. By summing up randomly drawn monthly expenditures we can determine whether or not an individual's expenditures would have exceeded the cost-sharing limit at the beginning of a given month, and we can then for each month compute the simulated fraction of individuals in a group who have not hit the cost-sharing limit. By pooling health care expenditures across months, the instrument is purged of differences in seasonal patterns across groups in the population. By pooling expenditures across years, we make sure that the variation in the instrumental variable across years is driven by changes in institutional rules, but not by how individuals respond to these changes. Simulated average prices for the same group and month differ across years because of changing cost-sharing limits, but not because of patients' responses to changing cost-sharing limits or because of different responses to prices in years with a deductible and in years with a no-claim refund.

In a different context, Currie and Gruber (1996a,b) and Cutler and Gruber (1996) use a similar approach to construct an instrument that is purged of potentially confounding factors

[22] Age groups are defined as age 0-18, 19-39, 40-64, and 65+. For the first four risk score deciles, we combine the two last age groups into one to ensure that there are enough observations per group.

Figure 5: Simulated fraction of individuals below cost-sharing limit by risk score



Notes: This figure shows the simulated fraction of individuals who have not hit the deductible. Average by quartile of the risk score, across years. Figure C.3 in the Online Appendix shows the actual fractions by risk score.

specific to states in the US: instead of using the number of individuals who are eligible for health insurance as the instrument, they use the simulated number, drawing from the distribution of characteristics at the national level and then applying state-specific eligibility rules to these hypothetical applicants. In that way, the instrument is purged of unobserved differences across states.

Figure 5 shows the average value of the instrument by month and by risk score quartiles. This figure illustrates the variation in average prices between groups that we use to identfy price effects.[23] For all enrollees, prices are one in January. For enrollees with high risk scores, average prices quickly decline over the following months and then approach zero. For enrollees with low risk scores, this decline is much slower and average prices are still close to one at the end of the year. In our empirical strategy we control for level differences in average prices between groups by including covariates x_{it}. The variation in prices that we use in our empirical approach comes from differences between groups in the *change* of average prices over the course of the year. These differences in the change of prices between groups are reflected in the different shapes of the curves depicted in Figure 5.

Formally, for the instrument z_{it} to be valid, we need it to be relevant and an exclusion

[23] Figure C.3 in the Online Appendix shows the actual fractions by risk score for 2007 and 2008. One can see that overall patterns are very similar, but that the actual fractions exhibit some seasonal variation.

restriction needs to hold. As for relevance, by design z_{it} is clearly related to p_{it}, and in Section 8.1 below, we document that the first stage F-statistic is very high. The exclusion restriction is that

$$\mathbb{E}\left[\varepsilon_{it}|z_{it}, x_{it}\right] = \mathbb{E}\left[\varepsilon_{it}|x_{it}\right],$$

which means that conditional on time dummies, age, gender, risk score, future price in January, and expenditures in the previous 3 months, the simulated fraction of individuals without cost-sharing is mean independent of ε_{it}.

In Section 6.1 we discussed three possible threats to identification if we estimate parameters in equation (1) by means of a linear regression: omitted variable bias, autocorrelation in error terms, and seasonal trends that differ between groups in the population. Our instrumental variable approach addresses these threats. In contrast to individual prices p_{it}, our instruemntal variable z_{it} is unlikely to be correlated with current or past individual error terms, since individual health expenditures are unlikely to influence average prices at the group level. In this way, our empirical approach addresses the first two threats to identification discussed in Section 6.1. Furthermore, by drawing expenditures for our simulation randomly across months our instrumental variables are purged of seasonal trends. This addresses the third threat to identification in Section 6.1.

In principle, a possible violation of the exclusion restriction comes from the fact that when estimating the empirical distribution function of monthly health care consumption, then we also use observations for individual i. Formally, the instrument is therefore a function of the error terms of individual i. However, the impact thereof will vanish in the limit, when the number of observations goes to infinity. We are not concerned about this because of the large number of observations. On average, there are more than 160,000 observations for each group in our estimation sample and the smallest group has 29,232 observations.[24]

7 Results

7.1 Main results

We now turn to our results. Our central hypothesis is that individuals react more strongly to financial incentives when they are framed as a deductible instead of a no-claim refund. This means that we expect the coefficient on the current price interacted with an indicator for the deductible regime to be more negative than the coefficient on the current price interacted with an indicator for the no-claim regime.

[24]In principle, one could construct the instrument at the individual level by estimating the empirical distribution function using data from all other individuals, but this would substantially increase the computational burden (by a factor N, where N is the number of individuals). We have experimented with this for a small number of individuals and have found that it did not noticeably affect the estimates of the empirical distribution function. Therefore, we have not implemented this for the full sample.

Table 2: Main results

	(1) expenditure	(2) has claim
(current price) x (noclaim regime)	-0.175***	-0.0526***
	(0.0461)	(0.00894)
(current price) x (deductible regime)	-0.362***	-0.0725***
	(0.0252)	(0.00477)
future price in January	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes
5 year age brackets	Yes	Yes
dummies decile risk score	Yes	Yes
year dummies	Yes	Yes
month dummies	Yes	Yes
number of observations	8,772,000	8,772,000
number of clusters	468	468
p-value equality current price coefficient	0.0000	0.0238

Notes: Instrumental variables estimates. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Our main results are shown in Table 2. The dependent variable in column (1) is the log of expenditures plus one.[25] Therefore, reported coefficients are (approximately) percentage changes.[26] We find the price effect to be a 17.5 percent decrease in expenditures under a no-claim refund and a 36.2 percent decrease under a deductible. This implies that under a no-claim refund individuals spend 17.5 percent less on health care in months before hitting the no-claim refund limit than in months after hitting the no-claim refund limit. Correspondingly, under a deductible individuals spend 36.2 percent less on health care in months before hitting the deductible limit than in months after hitting the deductible limit. While these estimates might seem large, our estimate for the effect of deductibles is similar and even somewhat smaller compared to the one of Brot-Goldberg et al. (2017), who find a price effect of deductibles on monthly health care expenditures of 42.2 percent for a sample of highly paid employees of a large American firm.

The last row of the table shows the p-value for a test of the null hypothesis that the two price effects are equal. This null hypothesis is rejected, in line with our main hypothesis that individuals react significantly stronger to prices under a deductible policy than under a no-claim refund policy.

In column (2), the dependent variable is an indicator for having a claim. Hence, reported coefficients show price effects at the extensive margin. The price effect is a 5.25 percentage point decrease in the probability to have a claim in a given month under the no-claim regime and a 7.22 percentage point decrease under the deductible. The difference is statistically different from zero at the 5 percent level.

7.2 Effect heterogeneity

Next, we characterize effect heterogeneity.[27] For this, we split the sample. We start with investigating whether price effects and the impact of framing differ by income. The price effect could differ between income groups if for example individuals with lower income react stronger to cost-sharing incentives than individuals with higher income, and the effect of framing could differ between income groups if they use different reference points. We divide individuals in 4 groups, according to the average income at the 6-digit neighborhood level, where quartile 1 refers to the group with the lowest income. As described in Section 4, a 6-digit postal code has on average 37 residents. Table 3 shows that results for each of the income groups are similar to the baseline results, and that there are no systematic differences across groups. So, individuals in poorer neighborhoods do not react stronger to cost-sharing incentives than individuals

[25] In Table 10 below we show that adding 10 or 0.1, respectively, leads to different point estimates but the same qualitative findings.

[26] Recall that we use the log of expenditures of 1 as the dependent variable. The exact percentage change of the dependent variable associated with a coefficient of -0.175 on the current price in the no-claim regime is $e^{-0.175} - 1 = -0.161$. In the following, we will abstract from such small differences.

[27] Here and in the following we present results for the effect on expenditures. Results for having a claim in a given month are presented in Online Appendix C.

Table 3: Results by income

	(1) quartile 1	(2) quartile 2	(3) quartile 3	(4) quartile 4
(current price) x (noclaim regime)	-0.161***	-0.218***	-0.193***	-0.170***
	(0.0519)	(0.0454)	(0.0487)	(0.0498)
(current price) x (deductible regime)	-0.376***	-0.363***	-0.384***	-0.355***
	(0.0300)	(0.0281)	(0.0287)	(0.0295)
future price in January	Yes	Yes	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes	Yes	Yes
5 year age brackets	Yes	Yes	Yes	Yes
dummies decile risk score	Yes	Yes	Yes	Yes
year dummies	Yes	Yes	Yes	Yes
month dummies	Yes	Yes	Yes	Yes
number of observations	2,052,132	2,206,728	2,083,332	2,084,652
number of clusters	467	467	468	468
p-value equality current price coefficient	0.0000	0.0004	0.0000	0.0001

Notes: Instrumental variables estimates for different subsamples defined by income quartile at the 6-digit neighborhood level. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table 4: Results by gender and age

	(1) female	(2) male	(3) age 19-64	(4) age 65+
(current price) x (noclaim regime)	-0.241***	-0.0436	-0.190***	-0.135**
	(0.0476)	(0.0448)	(0.0489)	(0.0614)
(current price) x (deductible regime)	-0.406***	-0.269***	-0.326***	-0.330***
	(0.0369)	(0.0278)	(0.0271)	(0.0408)
future price in January	Yes	Yes	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes	Yes	Yes
5 year age brackets	Yes	Yes	Yes	Yes
dummies decile risk score	Yes	Yes	Yes	Yes
year dummies	Yes	Yes	Yes	Yes
month dummies	Yes	Yes	Yes	Yes
number of observations	4,548,864	4,223,136	6,797,784	1,974,216
number of clusters	234	234	360	155
p-value equality current price coefficient	0.0000	0.0000	0.0047	0.0002

Notes: Instrumental variables estimates by age and gender, respectively. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

in richer neighborhoods. This finding is in line with evidence from the RAND health insurance experiment, where the effect of deductibles on total health care expenditures did also not significantly differ across income groups (Manning et al., 1987). Furthermore, within each group the price effect is significantly stronger under a deductible scheme than under a no-claim refund scheme. Also, the difference between the price effects under a deductible and a no-claim refund is similar across income groups.

Table 4 shows results by age and gender. Individuals in all age and gender groups react stronger to deductibles than to no-claim refunds. Moreover, price responses are similar for persons above age 65 and for persons below age 65. This is an interesting result since many previous studies on the effects of patient-cost sharing examine either the working age population or people above age 65, e.g. Medicare beneficiaries. In the Netherlands, individuals buy the same type of health insurance below and above age 65, and insurance coverage does not change discontinuously at age 65, as it does for instance in the U.S. (Card et al., 2008). However, we find differences in price effects on expenditures by gender. Women react stronger to cost-sharing incentives than men, both under a deductible and a no-claim refund.

A third way in which we split the sample is by risk score. The idea behind this is that we expect individuals with a higher risk score to react less strongly to price changes. About 80 percent of the individuals with an above-median risk score will exhaust the deductible or no-

Table 5: Results by risk score

	(1) low risk score	(2) high risk score
(current price) x (noclaim regime)	-0.446***	-0.266***
	(0.0882)	(0.0745)
(current price) x (deductible regime)	-0.615***	-0.446***
	(0.0943)	(0.0635)
future price in January	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes
5 year age brackets	Yes	Yes
dummies decile risk score	Yes	Yes
year dummies	Yes	Yes
month dummies	Yes	Yes
number of observations	4,335,084	4,436,916
number of clusters	198	270
p-value equality current price coefficient	0.0048	0.0000

Notes: Instrumental variables estimates by risk score. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

claim refund limit by the end of the year, while this is the case for only about 20-30 percent of the individuals with a below-median risk score. Therefore, one might expect below-median risk score individuals to react more strongly to hitting the cost-sharing limit, as it is *a priori* less likely that they do (see Section 3.1). Table 5 tentatively confirms this prediction. We find that the price effect for individuals with below-median risk scores is generally stronger than for individuals with above-median risk scores. However, respective confidence intervals intersect. Thus, the difference in price effects between risk score groups is not significantly different from zero.

7.3 Effect for selected types of care

So far, we have examined the effect of cost-sharing incentives on total health care expenditures, and we have shown that both no-claim refunds and deductibles reduce medical spending, but to different degrees. In order to make statements about welfare effects of cost-sharing schemes we would need to know the value of the care that patients reduce spending on. If patients reduce spending mostly on low value care then cost-sharing incentives can be welfare enhancing. If patients reduce spending mostly on high value care then cost-sharing schemes can be harmful. In general, it is difficult to classify treatments into high or low value care since the value of treatment depends on individual circumstances. However, there are some types of treatment that are generally considered high value care.[28]

Table 6 shows results for selected types of medication that are considered high value. Following Brot-Goldberg et al. (2017), we consider medication for diabetes type 1, diabetes type 2, high cholesterol level (statins), depression (anti-depressants), and high blood pressure (anti-hypertensives). These medications are commonly used, they provide major medical benefits, and they are typically not very expensive. This list includes both medications for which the absence of the medical benefits is salient and felt almost immediately (as for diabetes drugs and anti-depressants) and medications for which the absence of medical benefits is less salient (statins and anti-hypertensives). Statins and anti-hypertensives reduce the risk of heart diseases, but these benefits are not felt immediately.

The results in Table 6 show that the effects of deductibles and no-claim refunds on expenditures for selected high-value medications are small compared to the effects on total expenditures. For example, deductibles reduce monthly expenditures on diabetes type 1 medication by 0.4 percent while they reduce total monthly healthcare expenditures by 36.2 percent (see Table 2). Estimation coefficients are very precisely estimated, and some coefficients are statistically significant. But economically, these are small. Compared with the effects on total expenditures, the effects on expenditures for high value medication are negligible. Importantly, this holds for both no-claim refunds and deductibles.

[28] There are also some types of treatment that are considered potentially low value care, for instance imaging services. However, it is not feasible to identify individual procedures based on the information on bundled payments in our data.

Table 6: Results for selected types of care

	(1) diabetes type 1	(2) diabetes type 2	(3) statins	(4) anti-depressants	(5) anti-hypertensives
(current price) x (noclaim regime)	0.0142***	0.0137***	-0.0115	-0.00197	0.00227***
	(0.00303)	(0.00346)	(0.00795)	(0.00569)	(0.000590)
(current price) x (deductible regime)	-0.00437**	-0.00100	-0.0253***	0.00297	0.000534
	(0.00206)	(0.00258)	(0.00491)	(0.00242)	(0.000434)
future price in January	Yes	Yes	Yes	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes	Yes	Yes	Yes
5 year age brackets	Yes	Yes	Yes	Yes	Yes
dummies decile risk score	Yes	Yes	Yes	Yes	Yes
year dummies	Yes	Yes	Yes	Yes	Yes
month dummies	Yes	Yes	Yes	Yes	Yes
number of observations	8,772,000	8,772,000	8,772,000	8,772,000	8,772,000
number of clusters	468	468	468	468	468
p-value equality current price coefficient	0.0000	0.0000	0.0814	0.3890	0.0002

Notes: Instrumental variables estimates. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

7.4 The effect of framing on annual health care expendidures

Our main results so far are that no-claim refunds and deductibles reduce monthly health care expenditures by 17.5 percent and 36.2 percent, respectively. Yet, these coefficients are not indicative for the effects of no-claim refunds and deductibles on annual health care expenditures since cost-sharing incentives apply only for months in which cost-sharing limits are not yet exceeded. In this section we aim to compute the effect of no-claim refunds and deductibles on annual health care expenditures and on annual out-of-pocket spending.

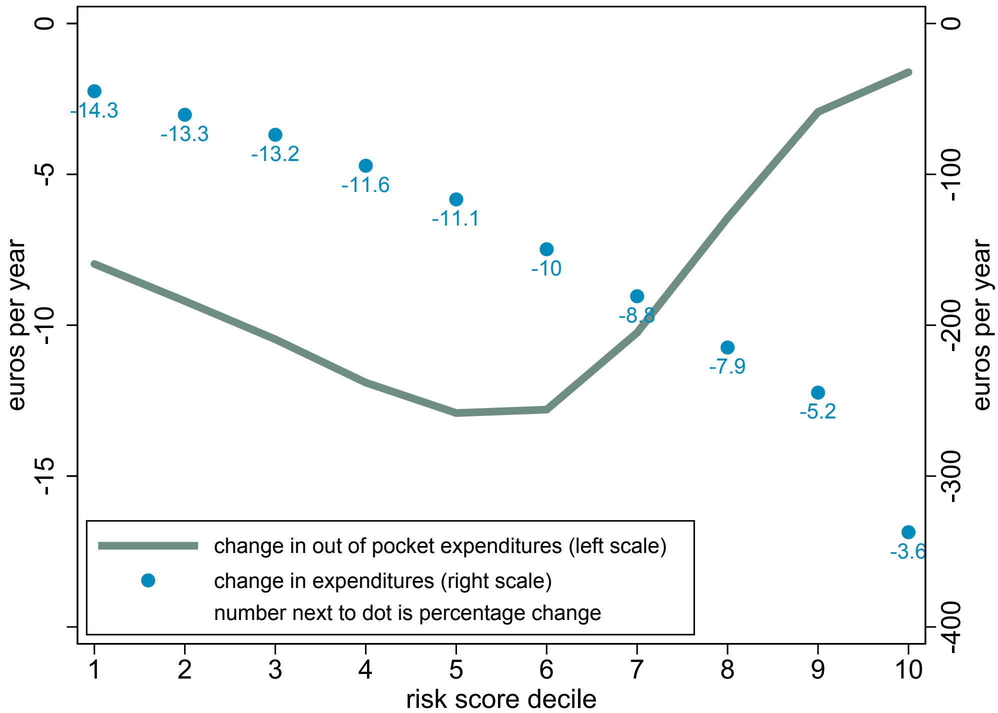
For this, we conduct a simulation study. In a first step, we calculate the residuals and estimate their empirical distribution function. Then, we take our estimation equation and a current price of 1 at the beginning of January as a starting point and simulate, for each individual in our data set, health care expenditures in January by calculating the predicted mean of the log of health care expenditures plus one, according to equation (1), and adding a draw from the empirical distribution function of the residual. This is then transformed into actual health care expenditures by applying the exponential function and subtracting 1. Next, we calculate simulated prices for all individuals at the beginning of February, again predict the mean of the log of health care spending plus one given those prices, and so on. We do so for 2015 and both, the case in which cost-sharing incentives are framed as a deductible, as it was the case in 2015, and for the case in which cost-sharing incentives are framed as a no-claim refund. For the latter, we use $\beta^{\text{no-claim refund}}$ instead of $\beta^{\text{deductible}}$.

Figure 6 shows the result. On average, annual health care expenditures are about 10 percent lower when cost-sharing incentives are framed as a deductible than when they are framed as a no-claim refund. In absolute terms, annual health care expenditures decrease more for higher risk score deciles. In relative terms, annual care expenditures decrease more for lower risk-score deciles. The underlying mechanism is that low risk-score individuals tend to exceed cost-sharing limits later in the year or not at all, and they are thus subject to cost-sharing incentives for a longer period than high risk-score individuals.

Turning to the effect of the framing of cost-sharing incentives on out-of-pocket spending, we find that it follows a U-shaped pattern. The effect is lowest for individuals with a high risk score. For individuals who exceed the cost-sharing limit under both cost-sharing schemes there is no difference in out-of-pocket spending between the two schemes. The effect is also small for individuals with the lowest risk-scores. These individuals have low health care expenditures and low out-of-pocket spending under both schemes. The effect of framing on out-of-pocket spending is largest for individuals in risk-score deciles 5 and 6.

Overall, it is striking that changing the framing of cost-sharing incentives leads to large effects on health care spending—a reduction of about 150 euros per year on average—, while having very modest effects—a reduction of about 10 euros on average—on out-of-pocket expenditures.

Figure 6: Effect of framing on yearly health care expenditures and out-of-pocket payments



Notes: This figure shows, for each risk score decile, the simulated effect of framing cost-sharing incentives as a deductible instead of a no-claim refund (dots). Numbers next to the dots are the percentage changes. Absolute values are given on the right axis. The solid line is the effect on out-of-pocket payments. The simulation is done for 2015.

Table 7: First stage

	(1) current price no claim	(2) current price deductible
simulated fraction group not hit		
interacted with noclaim regime	0.904***	-0.00714
	(0.0130)	(0.0126)
interacted with deductible regime	0.00297*	0.899***
	(0.00166)	(0.0112)
future price in January	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes
5 year age brackets	Yes	Yes
dummies decile risk score	Yes	Yes
year dummies	Yes	Yes
month dummies	Yes	Yes
number of observations	8,772,000	8,772,000
number of clusters	468	468
F-statistic excluded instruments	2542.52	6978.93

Notes: First stage estimates. We use data at the monthly level. Standard errors are clustered at the risk score decile-gender-age group-year level. $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

8 Validity of instrument and sensitivity analysis

In this section, we first assess whether the instruments we use are relevant. Then, we present results from a placebo test in order to assess the credibility of our empirical approach. After this, we ask the question whether the dependence of price effects on framing could also be explained by differences in the timing of payments. Finally, we present results for a number of robustness checks for alternative specifications of explanatory variables and outcome variables, and we examine whether our results are sensitive to changes in the sample composition and the basic health insurance plan over time. Here, we also ask whether end-of-the-year effects could explain our results.

8.1 Instrument relevance

One of the conditions for an instrument to be valid is that the instrument must be related to the endogenous variable conditional on controls. Table 7 shows that the simulated average price at the group level is highly predictive of the individual price p_{it}. The dependent variable in the first column is the individual price interacted with an indicator for the no claim regime, and the dependent variable in the second column is the individual price interacted with an indicator for

Table 8: Placebo test: no effect for individuals age 15-17

	(1) expenditure	(2) has claim
(current price) x (noclaim regime)	0.186	0.0107
	(0.120)	(0.0467)
(current price) x (deductible regime)	0.0898	0.0195
	(0.0629)	(0.0182)
future price in January	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes
5 year age brackets	Yes	Yes
dummies decile risk score	Yes	Yes
year dummies	Yes	Yes
month dummies	Yes	Yes
number of observations	359,736	359,736
number of clusters	171	171
p-value equality current price coefficient	0.4170	0.8557

Notes: Instrumental variables estimates for individuals who are between 15 and 17 years old. We compute risk scores and instruments separately for this group. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score-year level. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

the deductible regime. The respective coefficients on the simulated average prices interacted with the regime indicator are about 0.9. The relevant F-statistics, which are shown in the last row of the table, are extremely high. They are 2,543 and 6,979, respectively.

8.2 Placebo test

One way to assess the credibility of our empirical approach is to perform a placebo test. In the Netherlands children below the age of 18 are not subject to patient cost-sharing. For this reason, our main analysis was conducted for a sample of individuals at ages 19 and above.[29] If we apply our empirical approach to a sample of individuals between ages 15 and 17, we should not find any effect. And indeed, Table 8 shows that price effects are not significantly different from zero for this group and also not significantly different from one another. Thus, our empirical approach performs well in a setting where we know that the true parameters are zero.

[29] We did not include 18 year old individuals because we only observe the birth year and not the month of birth.

8.3 Can the difference in timing explain our results?

Our main finding is that individuals react stronger to deductibles than to no-claim refunds. So far, we interpreted our results as evidence for a strong effect of the framing of cost-sharing incentives on health care spending. Yet, deductibles and no-claim refunds differ not only in the framing of cost-sharing incentives, but also in the timing of payments. In general, deductibles have to be paid several weeks or months after the treatment, whereas the no-claim refund is only paid out at the end of the first quarter of the following year. This raises the question whether this difference in the timing of payments could explain our results.

In this section, we present empirical evidence that strongly speaks against this alternative explanation. In order to examine whether different responses can be explained by differences in the timing of payments, we re-do our analysis and take the delay in payments into account. For this, we assume a high, but still plausible annual discount rate of 10 percent, and we allow for compounding within the year. We further assume a delay of 3.5 months under the deductible regime, corresponding to the deductible payment being due at the end of the third month after the treatment. For the no-claim refund, we assume that the delay is 3.5 months in December, 4.5 months in November, and so on, because the no-claim refund is paid out at the end of the first quarter in the following year. Making these assumptions allows us to re-compute current prices and thus taking the difference in timing into account. In particular, we set the current price to $(1/(1+0.1))^{m/12}$ for payments m months in the future.

Results are presented in Table 9. They are very similar to our main results presented in Table 2. Thus, discounting payments at a 10 percent rate is not enough to explain our findings. In addition, we use a stylized model to compute which discount rate would be high enough to explain the difference in effects between deductibles and no-claim refunds that we find. Online Appendix B contains details of our calculations. We find that the annual discount rate would need to be higher than 300% to explain our results.

One possible explanation for this very high discount rate could be liquidity constraints. In our context, we believe that for institutional reasons liquidity constraints are unlikely to explain our results. The maximal amount an individual had to pay was 375 euros per year in 2015 (see Section 2), which is an amount that most Dutch persons can raise. In case this proves difficult then individuals can arrange payment in installments with the health insurer if they wish to do so. Moreover, as explained in Section 2.3, usually the payment is not immediately due at the time of treatment since the the treatment and the billing dates (to the insurance) do not coincide. Figure C.5 in the Online Appendix shows that the average delay between the treatment and billing date is about 60 days, and there will be an additional delay before the insurance will send an invoice to the individual.

Evidence that liquidity constrains are unlikely to explain the different effect of deductibles and no-claim refunds also comes from our analysis for different income groups at the 6-digit post code level. We would expect that liquidity constraints are more common for individuals

Table 9: Specification with discounted prices

	(1) expenditure	(2) has claim
(discounted current price) x (deductible regime)	-0.193***	-0.0580***
	(0.0510)	(0.00988)
(discounted current price) x (noclaim regime)	-0.372***	-0.0742***
	(0.0258)	(0.00488)
future price in January	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes
5 year age brackets	Yes	Yes
dummies decile risk score	Yes	Yes
year dummies	Yes	Yes
month dummies	Yes	Yes
number of observations	8,772,000	8,772,000
number of clusters	468	468
p-value equality current price coefficient	0.0002	0.0929

Notes: Instrumental variables estimates. We take the difference in timing into account by using the discounted current price instead of the actual current price. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

with lower incomes. However, in Table 3 we do not find that the price effects differ between income groups.

Taken together, our results strongly suggest that the difference in the timing of payments is not one of the main explanations for our finding that patients react stronger to deductibles than to no-claim refunds.

8.4 Robustness

In this subsection, we present a number of additional robustness checks. We first examine whether or not our estimation results are sensitive to different specifications of the covariates in x_{it}. These results are shown in Table 10. Recall that in our baseline specifications x_{it} includes variables for 5 year age brackets, gender, year and month dummies, indicators for the decile of the risk score, the future price in January, and the log of expenditure in the previous 3 months plus 1. In the first column of Table 10 we use a specification with a full set of risk score-year interactions and a full set of year-month interactions. The results are similar to the baseline results presented in Table 2, suggesting that a specification that is not fully-interacted in this way is general enough. For the second column we include 6 lags of the log of the respective monthly expenditure plus 1 instead of the log of the expenditure in the previous three months plus 1. Results are qualitatively similar as well. In the last column, we use the baseline specification but do not control for the expected price at the end of the year. This again leads to qualitatively similar results, but now the magnitude of the estimated effects is higher. Thus, we find that our main finding that the price effect is more negative under a deductible than under a no-claim refund is robust to using different specifications of explanatory variables.

Next, we examine whether or not our estimation results are sensitive to the specification of the outcome variable. Recall that up to now, we have used the log of expenditures plus 1. If we instead use the log of expenditures plus 0.1 or plus 10, then coefficients will be rescaled.[30] Table 11 shows that indeed, adding 0.1 and 10 instead of 1 leads to different estimates—as could be expected. But importantly, the main finding remains unchanged: the price effect is more negative under a deductible than under a no-claim refund. Table 11 also shows that estimation results remain essentially unchanged when we change the definition of health care expenditure and exclude expenditures on mental health care. Mental health care was included in the basic health insurance plan only since the year 2008.

After having assessed the robustness to using different specifications for the explanatory and dependent variables, respectively, we now investigate whether results are different for different

[30] For a simple example suppose that there is a claim with probability 0.4 under no cost-sharing and that the claim size is 50 conditional on there being a claim. Assume that the price effect of cost-sharing is a reduction in the probability that there is a claim by 10 percentage points and a reduction in the claim size by 50 percent once there is a claim. Then, the difference between the expected log of expenditure plus 1 under no cost-sharing and the expected log of expenditure plus 1 under cost-sharing is -0.33. If we use respectively 0.1 and 10 instead of 1, then we arrive at -0.42 and -0.23. This shows that the constant one adds will matter. The estimated effect will be tentatively be smaller in magnitude if we add a bigger number, at least in this example.

Table 10: Alternative specifications explanatory variables

	(1) flexible form	(2) distributed lag	(3) no future price
(current price) x (noclaim regime)	-0.200***	-0.206***	-0.552***
	(0.0609)	(0.0262)	(0.0694)
(current price) x (deductible regime)	-0.346***	-0.312***	-0.715***
	(0.0228)	(0.0211)	(0.0474)
future price in January	Yes	Yes	No
log exp. previous 3 months plus 1	Yes	No	Yes
5 year age brackets	Yes	Yes	Yes
dummies decile risk score	No	Yes	Yes
year dummies	No	Yes	Yes
month dummies	No	Yes	Yes
6 distributed lags log expenditure	No	Yes	No
risk score-year dummies	Yes	No	No
year-month dummies	Yes	No	No
number of observations	8,772,000	8,772,000	8,772,000
number of clusters	468	468	468
p-value equality current price coefficient	0.0326	0.0000	0.0037

Notes: Instrumental variables estimates. This table shows results using alternative specifications of the explanatory variables. In the first column we use a more flexible specification with risk score-year dummies and year-month dummies. The next column uses a distributed lag structure instead of the log of the expenditure in the previous 3 months. In the last column we do not condition on the future price in January. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

Table 11: Alternative specifications outcome variable

	(1) log(exp.+0.1)	(2) log(exp.+10)	(3) log(exp.) without mental health care	(4) has claim without mental health care
(current price) x (noclaim regime)	-0.296***	-0.0590**	-0.175***	-0.0526***
	(0.0646)	(0.0294)	(0.0461)	(0.00894)
(current price) x (deductible regime)	-0.528***	-0.202***	-0.362***	-0.0725***
	(0.0351)	(0.0162)	(0.0252)	(0.00477)
future price in January	Yes	Yes	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes	Yes	Yes
5 year age brackets	Yes	Yes	Yes	Yes
dummies decile risk score	Yes	Yes	Yes	Yes
year dummies	Yes	Yes	Yes	Yes
month dummies	Yes	Yes	Yes	Yes
number of observations	8,772,000	8,772,000	8,772,000	8,772,000
number of clusters	468	468	468	468
p-value equality current price coefficient	0.0001	0.0000	0.0000	0.0238

Notes: Instrumental variables estimates. We use data at the monthly level. Standard errors are clustered at the risk score decile-gender-age group-year level. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

definitions of our sample. Results are shown in Table 12. In column (1) we restrict the sample period to one year before and after the change from a no-claim refund to a deductible, by using data for the years 2007 and 2009 only. We have excluded data from the year 2008, because in 2008 we would use data from 2007 to construct control variables. By excluding data from 2008, we have a cleaner separation between the period in which the no-claim refund was in place and the period in which there was a deductible.[31] Results are similar to our main results for all years. In column (2) we show results for a balanced sample of individuals who are in our sample in all years from 2006 until 2015. In this way, we assess whether our results are sensitive to changes in the sample composition over time during our study period. Results for the balanced sample are similar to the baseline results for the full sample.

Our final robustness check concerns the shifting of health care expenditures across years (Cabral, 2016). For some types of health care, individuals can influence the timing of treatment. For example, they can choose whether they will receive a cataract surgery now or some months later. Individuals who have exceeded their cost-sharing limit in the current year have an incentive to shift treatments from the next year to the current year, because at the start of the new year patient cost-sharing limits will be reset. For those patients, the price of care in the current year is zero, but the price of care at the beginning of the next year is one. To see whether the shifting of care between years has an impact on our estimation results we estimate a model

[31] Figure 4 above was for 2007 and 2008. Figure C.4 in the Online Appendix is for 2007 and 2009 and corresponds more closely to the specification we use here.

Table 12: Alternative samples

	(1) only 2007 and 2009	(2) balanced sample
(current price) x (noclaim regime)	-0.233***	-0.101**
	(0.0533)	(0.0472)
(current price) x (deductible regime)	-0.334***	-0.343***
	(0.0470)	(0.0246)
future price in January	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes
5 year age brackets	Yes	Yes
dummies decile risk score	Yes	Yes
year dummies	Yes	Yes
month dummies	Yes	Yes
number of observations	1,787,376	3,855,276
number of clusters	104	468
p-value equality current price coefficient	0.0702	0.0000

Notes: Instrumental variables estimates. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table 13: End-of-the-year effects

	(1) December interaction	(2) drop December
(current price) x (noclaim regime)	-0.161***	-0.164***
	(0.0463)	(0.0479)
(current price) x (deductible regime)	-0.348***	-0.334***
	(0.0253)	(0.0252)
(current price) x (December)	-0.117***	
	(0.0147)	
future price in January	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes
5 year age brackets	Yes	Yes
dummies decile risk score	Yes	Yes
year dummies	Yes	Yes
month dummies	Yes	Yes
number of observations	8,772,000	8,041,000
number of clusters	468	468
p-value equality current price coefficient	0.0000	0.0002

Notes: Instrumental variables estimates. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

that in addition to the other covariates also includes an interaction term between the price p_{it} an an indicator for the month of December. The results presented in the first column of Table 13 show that this does not substantially change the coefficients for the price effects under the two cost-sharing schemes. Interestingly, however, individuals do seem to be more sensitive to the price of care in December than in other months, as one would expect. In the second column we do not use data for December and find results that are very similar to the ones presented in the first column and to our baseline estimates.

9 Conclusions

In this study we compare the effects of patient cost-sharing incentives on health care expenditures under a deductible policy and a no-claim refund policy. We show that patients react more strongly to deductibles than to no-claim refunds. This finding holds across income groups, for older and younger people, men and women, and for individuals with both high and low risk scores. Our results are robust to alternative specifications of both the outcome variable and the explanatory variables. In a placebo test we do not find significant effects if we apply our approach to 15-17 year olds who in the Netherlands are not subject to patient cost-sharing. Furthermore, descriptive evidence shows that the main result that individuals react stronger to deductibles than to no-claim refunds can also be seen in the raw data. Our preferred explanation for these results is that individuals are loss-averse and respond differently to both scheme since they perceive a deductible payment as a loss and a no-claim refund as a gain. We can exclude alternative explanations such as the different timing of payments under deductibles and no-claim refunds or the shifting of health care expenditures across years.

Our study adds to the growing evidence that patients' responses to cost-sharing incentives are subject to strong behavioral biases (Baicker et al., 2015). We contribute to the literature by showing that the framing of cost-sharing incentives has a strong effect on health care expenditures. In fact, we find that the framing of incentives can be quantitatively as important as the incentive itself.

An important remaining question is whether deductibles or no-claim refunds should be preferred from a welfare point of view. This question is difficult to answer. The original academic justification for patient cost-sharing incentives is to reduce moral hazard (Pauly, 1968). According to standard economic theory, individuals weigh the benefits and costs of treatment in their decisions. With full health insurance coverage the financial costs of health care utilization is zero, which can lead to moral hazard. Thus, reductions in medical spending as a result of patient cost-sharing should be seen as a welfare gain (Manning et al., 1987). Based on this reasoning, deductibles are preferable to no-claim refunds, because they lead to stronger reductions in medical spending.

Two concerns about this type of reasoning are that patients might be ignorant about the true benefits of medical treatment (Baicker et al., 2015) or that they might be financially constrained (Fels, 2017). Consequently, cost-sharing incentives might cause patients not only to forego low-value care, but also to forego care that is highly beneficial. In our study we find that the effect of both no-claim refunds and deductibles on 5 selected types of high value care is economically negligible. We also find that low income groups do not react stronger to cost-sharing incentives than high income groups. Thus, in our setting taking into account that patients might be uninformed about medical benefits or financially constrained does not change the conclusion that deductibles might be preferable to no-claim refunds, because they lead to stronger reductions in medical spending.

However, this conclusion could change if we allow for different utility weights of gains and losses under loss aversion. If individuals are loss-averse and use their financial situation after paying insurance premiums as reference point—as our results suggest—, then they might prefer no-claim refunds over deductibles (Johnson et al., 1993), even if overall health care costs are higher. This can have implications for the design of insurance plans. Some Dutch health insurers have recently started to offer insurance contracts that allow insurees to pre-pay their deductibles in monthly installments. For these contracts, individuals receive a refund at the end of the year if they have not exceeded their deductible limit, effectively turning the deductible into a no-claim refund. These schemes are popular among insurance holders.

Yet, in their calculation of insurance premiums Dutch insurers currently do not take into account that the different framing of cost-sharing incentives can lead to higher health care expenditures. According to our simulation results total annual health care expenditures are around 10 percent higher under a no-claim refund than under a deductible. Thus, offering the seemingly innocent option to pre-pay deductibles in monthly installments might cause losses for insurers.

Overall, designing cost-sharing schemes in health insurance contracts is a complex task. Many factors can influence patients' decision about health care use, and it is important to identify the most important factors. Our study shows that the framing of cost-sharing incentives is a factor of first-order importance.

References

Abaluck, J., J. Gruber, and A. Swanson (2015). Prescription drug use under Medicare Part D: A linear model of nonlinear budget sets. NBER Working Paper No. 20976. National Bureau of Economic Research, Boston, MA.

Baicker, K., S. Mullainathan, and J. Schwartzstein (2015). Behavioral hazard in health insurance. *Quarterly Journal of Economics 130*(4), 1623–1667.

Baucells, M., M. Weber, and F. Welfens (2011). Reference-point formation and updating. *Management Science 57*(3), 506–519.

Brot-Goldberg, Z. C., A. Chandra, B. R. Handel, and J. T. Kolstad (2017). What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics. *Quarterly Journal of Economics 132*(3), 1261–1318.

Brown, J. R., A. Kapteyn, and O. S. Mitchell (2016). Framing and claiming: How information-framing affects expected Social Security claiming behavior. *Journal of Risk and Insurance 83*(1), 139–162.

Cabral, M. (2016). Claim timing and ex post adverse selection. *Review of Economic Studies 84*(1), 1–44.

Card, D., C. Dobkin, and N. Maestas (2008). The impact of nearly universal insurance coverage on health care utilization: Evidence from Medicare. *American Economic Review 98*(5), 2242–2258.

Chandra, A., J. Gruber, and R. McKnight (2010). Patient cost-sharing and hospitalization offsets in the elderly. *American Economic Review 100*(1), 193.

Currie, J. and J. Gruber (1996a). Health insurance eligibility, utilization of medical care, and child health. *Quarterly Journal of Economics 111*(2), 431–466.

Currie, J. and J. Gruber (1996b). Saving babies: the efficacy and cost of recent changes in the medicaid eligibility of pregnant women. *Journal of Political Economy 104*(6), 1263–1296.

Cutler, D. M. and J. Gruber (1996). Does public insurance crowd out private insurance? *Quarterly Journal of Economics 111*(2), 391–430.

Dalton, C. M., G. Gowrisankaran, and R. Town (2017). Salience, myopia, and complex dynamic incentives: Evidence from Medicare Part D. NBER Working Paper No. 21104. National Bureau of Economic Research, Boston, MA.

Duflo, E. (2017). The economist as plumber. NBER Working Paper No. 23213. National Bureau of Economic Research, Boston, MA.

Ecorys (2011). Evaluatie naar het verplicht eigen risico. Ecorys, Rotterdam, The Netherlands.

Einav, L., A. Finkelstein, S. P. Ryan, P. Schrimpf, and M. R. Cullen (2013). Selection on moral hazard in health insurance. *American Economic Review 103*(1), 178–219.

Einav, L., A. Finkelstein, P. Schrimpf, et al. (2015). The response of drug expenditure to nonlinear contract design: Evidence from Medicare Part D. *Quarterly Journal of Economics 130*(2), 841–899.

Ellis, R. P. (1986). Rational behavior in the presence of coverage ceilings and deductibles. *RAND Journal of Economics 17*(2), 158–175.

Ellis, R. P., B. Martins, and W. Zhu (2017). Health care demand elasticities by type of service. *Journal of Health Economics 55*, 232–243.

Fels, M. (2017). Incentivizing efficient utilization without reducing access: The case against cost-sharing in insurance. Karlsruhe Institute of Technology Working Paper Series No 105. Karlsruhe, Germany.

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, K. Baicker, Oregon Health Study Group, and others (2012). The Oregon health insurance experiment: Evidence from the first year. *Quarterly Journal of Economics 127*(3), 1057–1106.

Hasaart, F. (2011). *Incentives in the Diagnosis Treatment Combination payment system for specialist medical care: A study about behavioral responses of medical specialists and hospitals in the Netherlands*. Ph. D. thesis, Universiteit Maastricht, Maastricht, The Netherlands.

Hayen, A. P., T. J. Klein, M. Salm, and S. Upadhyay (2017). How forward-looking are individuals? Working Paper, Tilburg University. Tilburg, The Netherlands.

Hayen, A. P., M. J. van den Berg, B. R. Meijboom, J. N. Struijs, and G. P. Westert (2015). Incorporating shared savings programs into primary care: From theory to practice. *BMC Health Services Research 15*(580), 1–15.

Johnson, E. J., J. Hershey, J. Meszaros, and H. Kunreuther (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty 7*(1), 35–51.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–292.

Keeler, E. B. and J. E. Rolph (1988). The demand for episodes of treatment in the health insurance experiment. *Journal of Health Economics 7*(4), 337–367.

Krabbe-Alkemade, Y., T. Groot, and M. Lindeboom (2016). Competition in the Dutch hospital sector: An analysis of health care volume and cost. *European Journal of Health Economics*, 1–15.

Lambregts, T. R. and R. C. van Vliet (2017). The impact of copayments on mental healthcare utilization: A natural experiment. *European Journal of Health Economics*, 1–10.

Manning, W., J. Newhouse, N. Duan, E. Keeler, A. Leibowitz, and M. Marquis (1987). Health insurance and the demand for medical care: Evidence from a randomized experiment. *American Economic Review 77*(3), 251–277.

Netherlands Institute for Social Research (2006). Statusscores postcodegebieden. https://www.scp.nl/Onderzoek/Lopend_onderzoek/A_Z_alle_lopende_onderzoeken/Statusscores. The Hague, The Netherlands.

NZA (2012). Marktscan zorgverzekeringsmarkt 2012. Dutch Healthcare Authority NZA, Utrecht, The Netherlands.

OECD Indicators (2015). Health at a glance 2015. OECD Publishing. Paris, France.

Pauly, M. V. (1968). The economics of moral hazard: Comment. *American Economic Review 58*(3), 531–537.

Reitsma-van Rooijen, M. and J. d. Jong (2009). Compensatie eigen risico is nog onbekend. *Zorg en Financiering 8*(6), 34.

Remmerswaal, M., J. Boone, M. Bijlsma, R. Douven, et al. (2017). Cost-sharing design matters: A comparison of the rebate and deductible in healthcare. CPB Discussion Paper 367, CPB Netherlands Bureau for Economic Policy Analysis, The Hague, The Netherlands.

Remmerswaal, M., R. Douven, and P. Besseling (2015). Toelichting op de effecten van eigen betalingen in de zorg. CPB Achtergronddocument, CPB Netherlands Bureau for Economic Policy Analysis, The Hague, The Netherlands.

Schäfer, W., M. Kroneman, W. Boerma, M. Berg, G. Westert, W. Devillé, E. v. Ginneken, et al. (2010, 1). The Netherlands: Health system review. *Health Systems in Transition 12*(1). European Observatory on Health Systems and Policies.

Schmitz, H. and N. R. Ziebarth (2016). Does price framing affect the consumer price sensitivity of health plan choice? *Journal of Human Resources 52*(1), 89–128.

Statistics Netherlands (2012). Inhoud kerncijfers postcodegebieden 2008-2010. https://www.cbs.nl/nl-nl/maatwerk/2012/01/kerncijfers-postcodegebieden-2008-2010, CBS Statistics Netherlands, The Hague, The Netherlands.

Van de Ven, W. P. and F. T. Schut (2009). Managed competition in the Netherlands: Still work-in-progress. *Health Economics 18*(3), 253–255.

van Winsen, K. P. M., R. C. van Kleef, and W. P. van de Ven (2016). Potential determinants of deductible uptake in health insurance: How to increase uptake in the Netherlands? *European Journal of Health Economics 17*(9), 1059–1072.

World Health Organization (2006). The anatomical therapeutic chemical classification system with defined daily doses (atc/ddd). http://www.who.int/classifications/atcddd/en/.

Online Appendix

This Online Appendix contains additional details on the data, the derivation of the implied discount rate, and additional tables and figures.

A Additional details on the data

A.1 Services that are subject to cost-sharing

Constructing our main independent variables and the price requires that we know which services are subject to cost-sharing, as some services are exempted from cost-sharing. Such a list was not available to us. We therefore proxied for this using data on actual deductible payments made and information on the overarching care types that are exempt from the deductible, as specified by law. The types of care that are exempted include obstetric and maternal care, care generally provided by General Practitioners (GPs) and participation under a chronic care program.

More specifically, our approach for determining coverage consisted of the following steps (in chronological order):

1. Every specialist service was defined as counting towards the deductible;
2. Services provided by registered obstetricians were defined as not counting towards the deductible;
3. Specialist services were not defined as counting towards the deductible when they were received only by women (90%, to account for administrative errors) and,
   (a) for which in fewer than 1% of the claims a deductible payment was requested, for services that we observed over a 1000 times in the data or
   (b) for which in fewer than 10% of the claims a deductible payment was requested for services that observed less than a 100 times in the data;
4. For any remaining services, we determined coverage along the lines of 3a and 3b. That is, we let coverage fully depend on the percentage of claims in which actual deductible payments were made;
5. Coverage under the no-claim regime was deduced from 2008 deductible data.

Using the percentage cut-off points in 3a and 3b, we take account of administrative errors. We used a more conservative cut-off points for services that were billed less frequently.

A.2 Risk scores

To calculate the risk score, we regress total health care costs for a given year on a gender dummy, fully interacted with a third order polynomial in age, indicators for the decile of costs in the previous year, indicators for chronic conditions, indicators for characteristics at the 6-digit neighborhood level. For the latter we conduct a median split for income, the fraction non-western immigrants, and the CBS socioeconomic status score, respectively. We also include dummies for missing characteristics at the neighborhood level. Finally, we obtain fitted value and divide them by the average.

B Implied discount rate

In our empirical analysis, we have obtained estimates of price effects, separately for the no-claim and deductible regime. Our interpretation of the difference between the estimated effects is that the difference arises due to framing. In principle, an alternative explanation is that individuals react stronger to deductibles because then they have to pay earlier, as the no-claim refund is only paid out at the beginning of the following year. Here, we conduct a simple back-of-the envelope calculation with the goal to determine what the individual discount factor must be in order to generate this effect. We find that it would have to be unreasonable low in order to generate this effect, supporting the interpretation of the effects as being due to framing.

For this, we use a framework with a quadratic utility function as in Einav et al. (2013). Denote health care needs in month t by λ_t, health care consumption by m_t, income by y_t, and the implications of cost-sharing by $\delta^\tau p_t$. δ is the monthly discount factor so that δ^τ is the discount factor associated with payments τ months later, when the "current price" is p_t. The flow utility in period t depends on health care consumption,

$$u(m_t) = (m_t - \lambda_t) - \frac{1}{2\omega}(m_t - \lambda)^2 + y - \delta^\tau p_t m_t.$$

Differentiating with respect to m_t allows us to solve for the optimal consumption in t,

$$m_t^*(p_t) = \lambda_t + \omega(1 - \delta^\tau p_t).$$

When individuals have to fully and immediately pay out-of-pocket, we have $\delta^\tau p_t = 1$. Then, optimal health care consumption is equal to health care needs. In contrast, after hitting the cost-sharing limit, $\delta^\tau p_t = 0$ so that health care consumption is equal to health care needs plus ω. For that reason, ω can be interpreted as *ex post* moral hazard.

The effects reported in this paper are percentage changes in health care consumption, where the base is consumption when the current price is equal to zero. In terms of this model, this is

$$\frac{m_t^*(1) - m_t^*(0)}{m_t^*(0)} = \frac{-\delta^\tau p_t}{\lambda_t + \omega}.$$

Different regimes have different discount factors because they have different delays for the payments. For a rough approximation we assume that the average treatment date is after the first half of the year, payment under the deductible regime occurs on average three months later, and payment under the payback rebate occurs at the end of the first quarter of the following year. This means that the difference in timing is 6 months on average.

Assuming that health care needs and moral hazard effects stay the same, the ratio of coefficients obtained in the main analysis is

$$\frac{-0.175}{-0.362} = \frac{\delta^9}{\delta^3},$$

so that the yearly discount factor would have to be

$$\delta^{12} = 0.234.$$

This corresponds to the arguably unrealistically high discount rate of

$$\frac{1}{0.234} - 1 = 328\%.$$

C Additional Tables and Figures

Table C.1: Summary statistics over time

	mean
care consumption	
2006	1368.50
2007	1478.81
2008	1585.28
2009	1720.32
2010	1824.06
2011	1944.91
2012	1987.26
2013	2186.48
2014	2285.09
2015	2583.54
hit cost-sharing limit	
2006	0.53
2007	0.57
2008	0.63
2009	0.64
2010	0.65
2011	0.65
2012	0.62
2013	0.56
2014	0.56
2015	0.56
number individuals	35,697

Notes: This table shows summary statistics for all individuals who were in the sample in January 2006. Care consumption is the consumption of care that falls under the no-claim refund policy or the deductible. Hit deductible is one for a person in a given year if that person exceeded the deductible or the payback limit in a given year.

Table C.2: Results for has claim by income

	(1) quartile 1	(2) quartile 2	(3) quartile 3	(4) quartile 4
(current price) x (noclaim regime)	-0.0571***	-0.0576***	-0.0548***	-0.0466***
	(0.00967)	(0.00932)	(0.00992)	(0.0106)
(current price) x (deductible regime)	-0.0724***	-0.0729***	-0.0802***	-0.0793***
	(0.00576)	(0.00526)	(0.00555)	(0.00587)
future price in January	Yes	Yes	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes	Yes	Yes
5 year age brackets	Yes	Yes	Yes	Yes
dummies decile risk score	Yes	Yes	Yes	Yes
year dummies	Yes	Yes	Yes	Yes
month dummies	Yes	Yes	Yes	Yes
number of observations	2,052,132	2,206,728	2,083,332	2,084,652
number of clusters	467	467	468	468
p-value equality current price coefficient	0.1139	0.0951	0.0072	0.0014

Notes: Instrumental variables estimates for different subsamples defined by income quartile at the 6-digit neighborhood level. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

Table C.3: Results for has claim by gender and age

	(1) female	(2) male	(3) age 19-64	(4) age 65+
(current price) x (noclaim regime)	-0.0572***	-0.0239***	-0.0494***	-0.0734***
	(0.00870)	(0.00900)	(0.0102)	(0.0153)
(current price) x (deductible regime)	-0.0689***	-0.0547***	-0.0722***	-0.0830***
	(0.00644)	(0.00472)	(0.00525)	(0.0104)
future price in January	Yes	Yes	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes	Yes	Yes
5 year age brackets	Yes	Yes	Yes	Yes
dummies decile risk score	Yes	Yes	Yes	Yes
year dummies	Yes	Yes	Yes	Yes
month dummies	Yes	Yes	Yes	Yes
number of observations	4,548,864	4,223,136	6,797,784	1,974,216
number of clusters	234	234	360	155
p-value equality current price coefficient	0.1204	0.0004	0.0249	0.4616

Notes: Instrumental variables estimates by age and gender, respectively. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

Table C.4: Results for has claim by risk score

	(1) low risk score	(2) high risk score
(current price) x (noclaim regime)	-0.126***	-0.0861***
	(0.0259)	(0.0145)
(current price) x (deductible regime)	-0.162***	-0.101***
	(0.0256)	(0.0125)
future price in January	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes
5 year age brackets	Yes	Yes
dummies decile risk score	Yes	Yes
year dummies	Yes	Yes
month dummies	Yes	Yes
number of observations	4,335,084	4,436,916
number of clusters	198	270
p-value equality current price coefficient	0.0565	0.1182

Notes: Instrumental variables estimates by risk score. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $* \; p < 0.10$, $** \; p < 0.05$, $*** \; p < 0.01$.

Table C.5: Results for has claim and selected types of care

	(1) diabetes type 1	(2) diabetes type 2	(3) statins	(4) anti-depressants	(5) anti-hypertensives
(current price) x (noclaim regime)	0.00316***	0.00901***	0.0322***	0.00775***	0.00104***
	(0.000651)	(0.00185)	(0.00505)	(0.00185)	(0.000193)
(current price) x (deductible regime)	-0.000804*	-0.00550***	-0.0210***	-0.00764***	0.0000905
	(0.000438)	(0.000996)	(0.00229)	(0.000987)	(0.000135)
future price in January	Yes	Yes	Yes	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes	Yes	Yes	Yes
5 year age brackets	Yes	Yes	Yes	Yes	Yes
dummies decile risk score	Yes	Yes	Yes	Yes	Yes
year dummies	Yes	Yes	Yes	Yes	Yes
month dummies	Yes	Yes	Yes	Yes	Yes
number of observations	8,772,000	8,772,000	8,772,000	8,772,000	8,772,000
number of clusters	468	468	468	468	468
p-value equality current price coefficient	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: Instrumental variables estimates. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table C.6: Results for has claim and alternative specifications explanatory variables

	(1) flexible form	(2) distributed lag	(3) no future price
(current price) x (noclaim regime)	-0.0485***	-0.0570***	-0.137***
	(0.0107)	(0.00657)	(0.0146)
(current price) x (deductible regime)	-0.0663***	-0.0616***	-0.152***
	(0.00413)	(0.00452)	(0.00974)
future price in January	Yes	Yes	No
log exp. previous 3 months plus 1	Yes	No	Yes
5 year age brackets	Yes	Yes	Yes
dummies decile risk score	No	Yes	Yes
year dummies	No	Yes	Yes
month dummies	No	Yes	Yes
6 distributed lags log expenditure	No	Yes	No
risk score-year dummies	Yes	No	No
year-month dummies	Yes	No	No
number of observations	8,772,000	8,772,000	8,772,000
number of clusters	468	468	468
p-value equality current price coefficient	0.1247	0.4304	0.2407

Notes: Instrumental variables estimates. This table shows results using alternative specifications of the explanatory variables. In the first column we use a more flexible specification with risk score-year dummies and year-month dummies. The next column uses a distributed lag structure instead of the log of the expenditure in the previous 3 months. In the last column we do not condition on the future price in January. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table C.7: Results for has claim and additional robustness checks

	(1) only 2007 and 2009	(2) balanced sample
(current price) x (noclaim regime)	-0.0458***	-0.0481***
	(0.00893)	(0.00882)
(current price) x (deductible regime)	-0.0701***	-0.0739***
	(0.00856)	(0.00496)
future price in January	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes
5 year age brackets	Yes	Yes
dummies decile risk score	Yes	Yes
year dummies	Yes	Yes
month dummies	Yes	Yes
number of observations	1,787,376	3,855,276
number of clusters	104	468
p-value equality current price coefficient	0.0121	0.0027

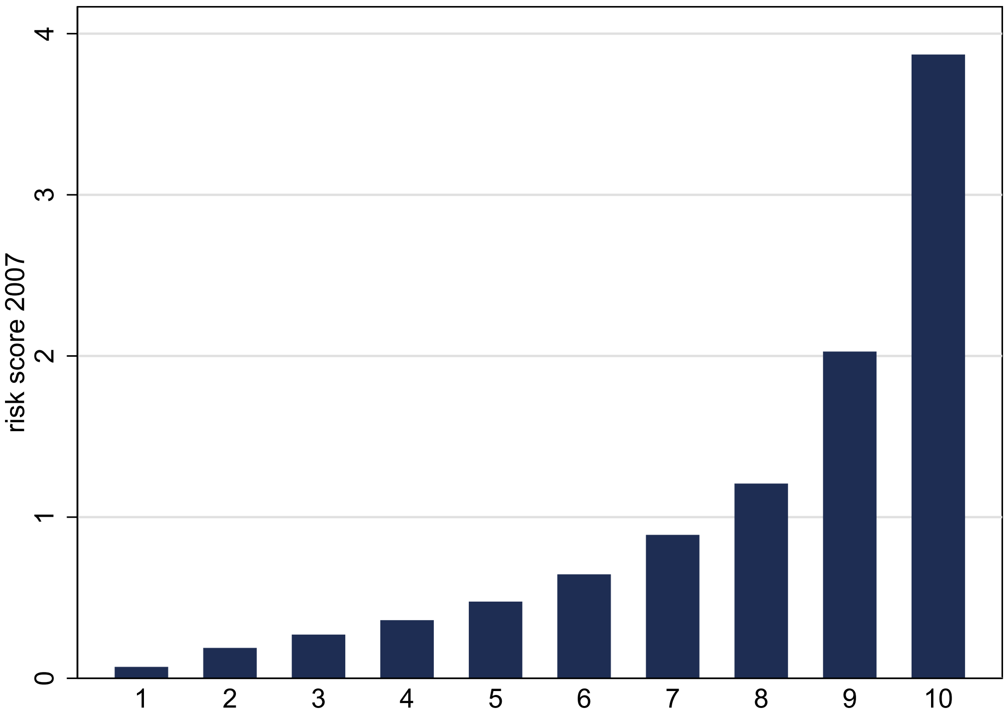
Notes: Instrumental variables estimates. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table C.8: End-of-the-year effects for has claim

	(1) December interaction	(2) drop December
(current price) x (noclaim regime)	-0.0501***	-0.0479***
	(0.00897)	(0.00919)
(current price) x (deductible regime)	-0.0700***	-0.0648***
	(0.00473)	(0.00455)
(current price) x (December)	-0.0202***	
	(0.00247)	
future price in January	Yes	Yes
log exp. previous 3 months plus 1	Yes	Yes
5 year age brackets	Yes	Yes
dummies decile risk score	Yes	Yes
year dummies	Yes	Yes
month dummies	Yes	Yes
number of observations	8,772,000	8,041,000
number of clusters	468	468
p-value equality current price coefficient	0.0242	0.0666

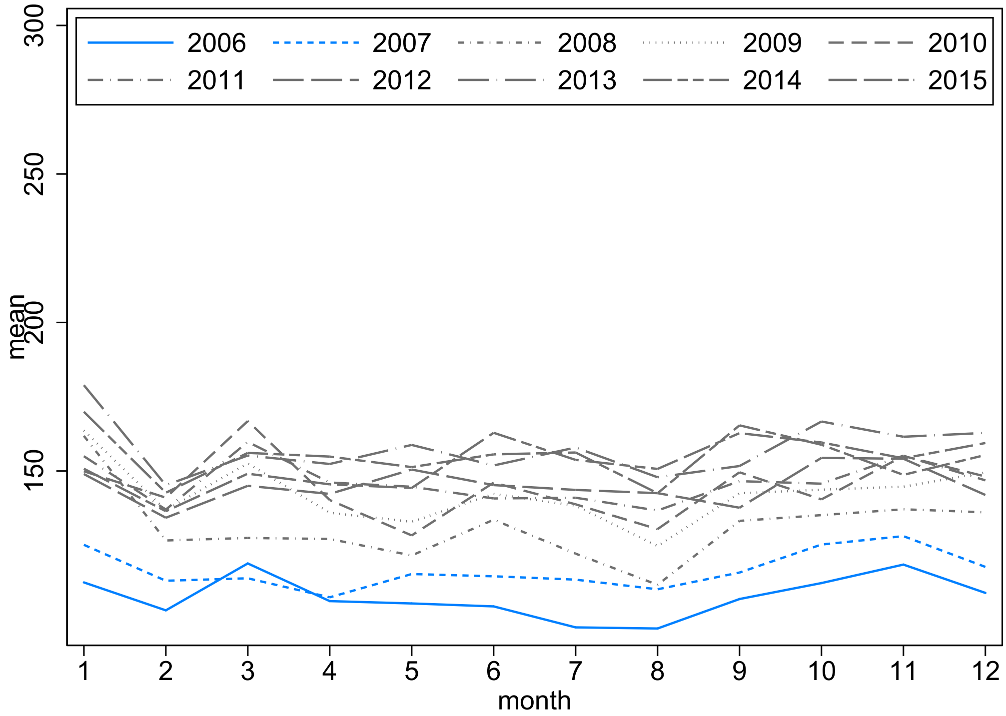
Notes: Instrumental variables estimates. We use data at the monthly level. Expenditures are measured as the log of one plus the actual amount. Standard errors are clustered at the risk score decile-gender-age group-year level. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Figure C.1: Risk score deciles



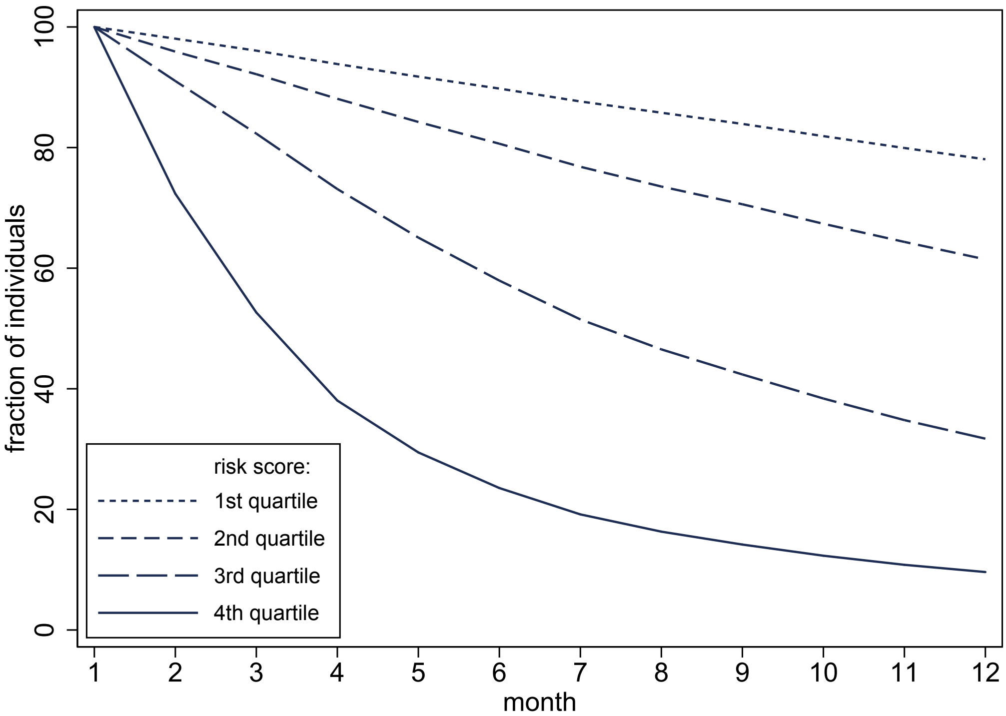
Notes: This figure shows risk score deciles. See Appendix A.2 for details on how the risk score was calculated.

Figure C.2: Care consumption over time



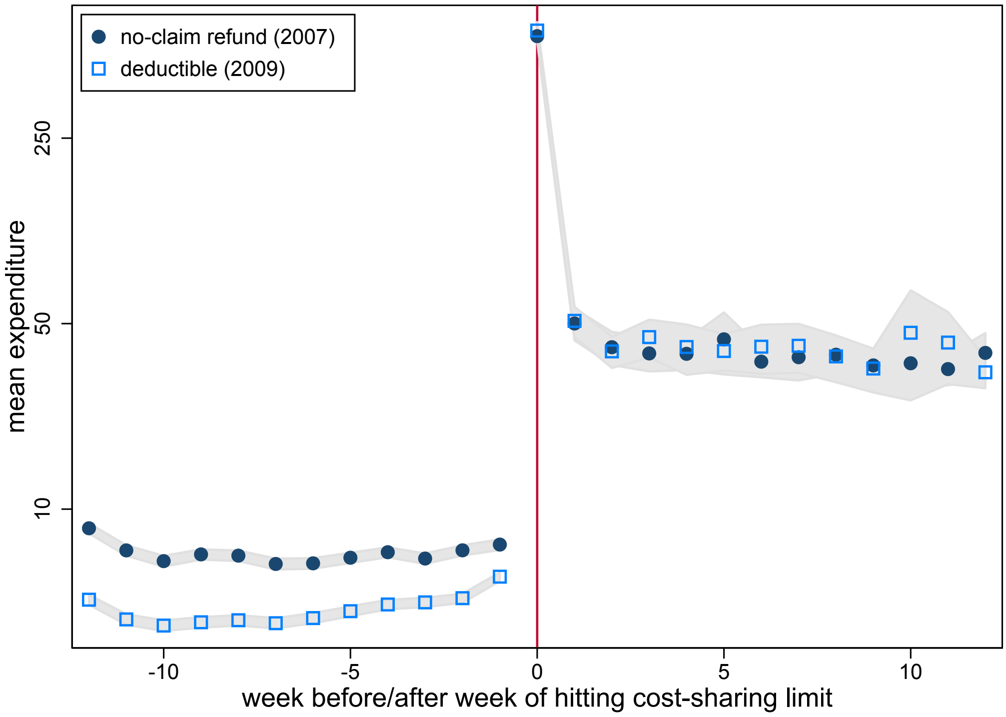
Notes: This figure shows care consumption, by month and for all years between 2006 and 2015. Calculated for the full unbalanced panel.

Figure C.3: Fraction of individuals below deductible by risk score



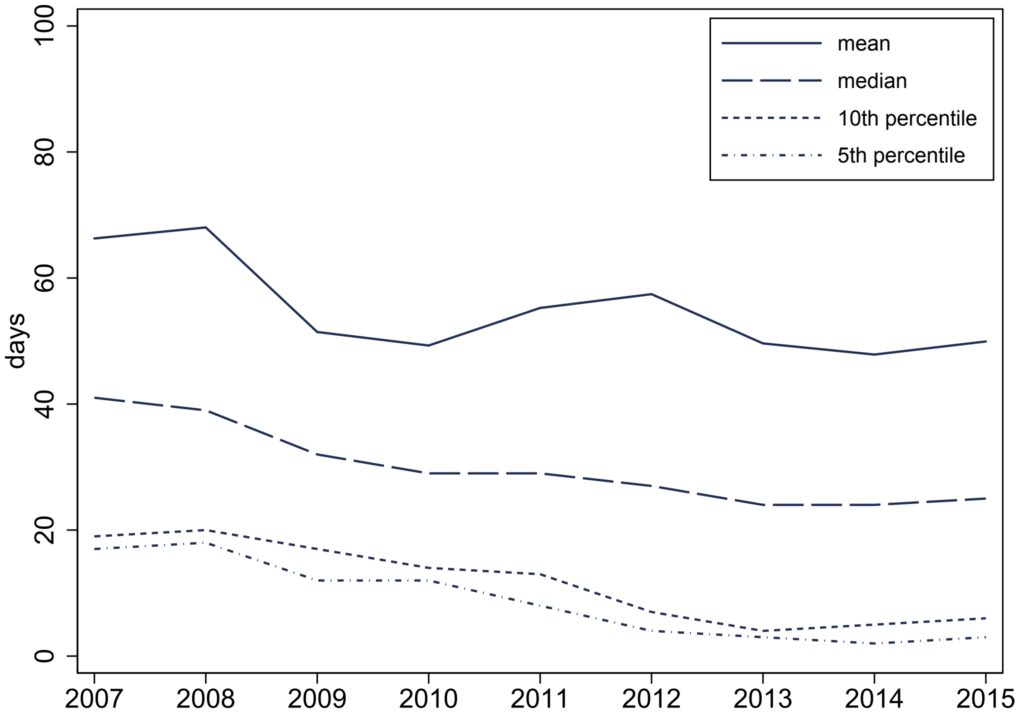
Notes: This figure shows the fraction of individuals who have not hit the deductible, by month and by quartile of the risk score, pooled over the years. Figure 3 shows the fractions by year.

Figure C.4: Care consumption around week in which cost-sharing ends



Notes: This figure shows average health care expenditures in weeks before and after exceeding the no-claim refund limit or the deductible in 2007 and 2009, respectively. Cost-sharing incentives were framed as a no-claim refund in 2007 and as a deductible in 2009. Figure 4 in the main text is for 2007 and 2008.

Figure C.5: Delay between treatment and billing date



Notes: This figure shows how the delay between treatment and billing date changed over time. Calculated using all claims in our data.