A Claim a Day Keeps the Doctor Away? Premium Refunds and Forward Looking Behaviour in the German Private Health Insurance Market*

Daniel Avdic[a,b], Simon Decker[a,b,c], and Martin Karlsson[a,b]

[a]CINCH – Health Economics Research Center
[b]University of Duisburg-Essen
[c]Ruhr Graduate School in Economics (RGS Econon)

May 2017

Preliminary and incomplete
Please do not circulate without permission

Abstract

Premium refunds, a payback agreement by the insurer contingent on a claim-free calendar year by the insured, is a common feature in the German private health insurance market but so far the evidence on its effects have been scarce. We study how such refunds impacts individual claiming behavior using rich administrative claims data from a large German health insurance company and an insurer policy that unexpectedly increased the refund size of certain plans. We furthermore suggest a novel method to decompose the overall effect on claims into an intensive, extensive and an automatic component. Our findings show that individuals reacted to the changed incentives by reducing their claims on both behavioral margins. Supplementary analyses suggest that these reductions led to increased health expenditures in later years.

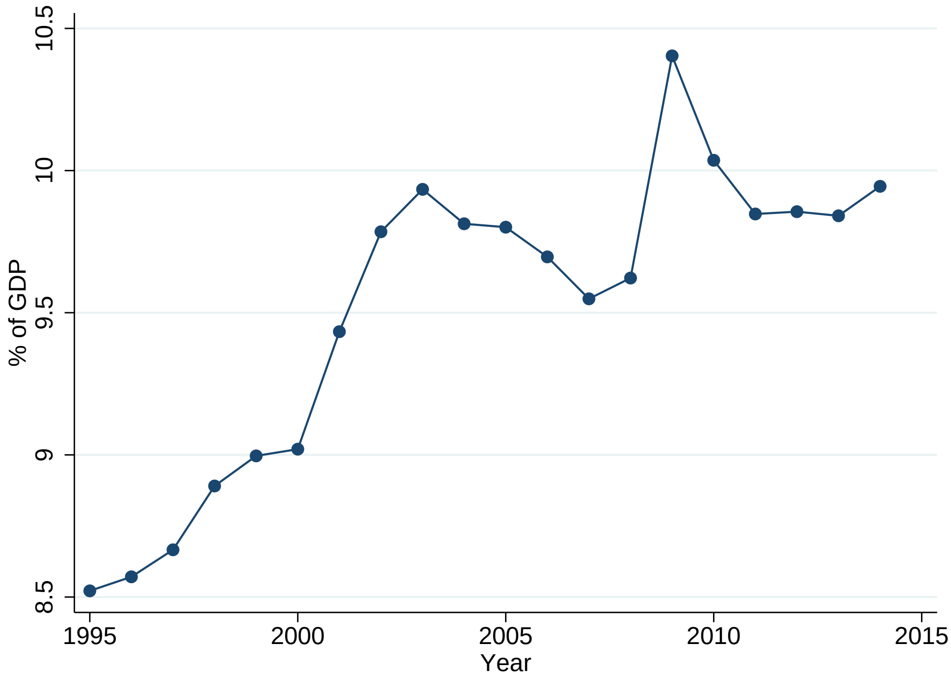
JEL-codes: D82, G22, I13
Keywords: premium refund, moral hazard, health insurance

*Correspondence to: Daniel Avdic, CINCH Health Economics Research Center, Room WST-C.14.12b, Weststadttürme Berliner Platz 6-8, 45127 Essen, +49(201)-183-6326, daniel.avdic@uni-due.de. We are grateful for valuable comments from John Buckell, Burkhard Hehenkamp, Hendrik Jürges, Jonathan Kolstad, Jing Lao, Giuseppe Moscelli, Robert Nuscheler, Carol Propper, Ieva Sriubaite, Johann Han, Vincent Pohl, Nicolas Ziebarth and Peter Zweifel and seminar participants in Barcelona, Bergen, Essen, Heidelberg, Hohenheim, Munich and Sapporo. All remaining errors are our own.

1 Introduction

Health care expenditures have been steadily increasing around the world over the last decades (see, e.g., Figure 1). This development has put pressure on public finances in many countries and cost containment strategies have therefore become prioritized by policy makers. One area of particular interest has been the health insurance market in which adverse selection and moral hazard are likely to be responsible for severe inefficiencies. The former implies that greater coverage rates will be chosen by individuals with poorer health, causing prices to be too high insurance coverage to be too low. The latter implies that individuals, once insured, consume more care than would be optimal due to cost-sharing.

FIGURE 1.
Global health expenditures, total (% of GDP)



NOTE.— The World Bank Global Health Expenditure database.

Previous evidence (cf., Carlin and Town, 2010; Einav *et al.*, 2012; Einav and Finkelstein, 2011; Bundorf *et al.*, 2012) have shown that adverse selection seem to be accountable for a relatively small share of the overall welfare costs associated with health insurance market failures. The main explanation seem to be that the demand for higher coverage seem to be quite inelastic due to inertia of insurance choice. Moral hazard, the elasticity of health care expenditures with respect to its out-of-pocket price, on the other hand, has been found to be a much more important factor in explaining variation in health care expenditures. Early empirical evidence from the well-known RAND health insurance experiment found a price elasticity of about –0.2 (Manning *et al.*, 1987; Keeler and Rolph, 1988). A recent reexamination by Aron-Dine *et al.* (2013) confirms this result. These findings thus imply that higher cost-sharing reduces medical spending and may reduce moral hazard.

One limitation with the evidence from the RAND experiment is that it assumes linear insurance contracts and therefore constant prices. A more recent line of research have instead discussed the implications of dynamic incentives due to non-linearities in health insurance plans arising from various types of co-payments associated with utilization (cf., Aron-Dine *et al.*, 2015; Dalton *et al.*, 2015; Einav *et al.*, 2015; Brot-Goldberg *et al.*, 2015). This distinction is crucial for several reasons; first, most contracts, including the choices included in the RAND experiment, have non-linear segments due to co-payments in the form of deductibles, reimbursement caps and other contract-specific features. Second, non-linear pricing implies that it is not clear which price(s) individuals will respond to for their choice of utilization. Assuming that individuals are forward-looking, not only the current ("spot") price matters, but also the expected future price will be important in the agent's decision to utilize health care.

In this paper we contribute to the literature on consumer behavior in health insurance markets by studying a largely unexplored tool used by health insurance providers to control costs; premium refunds. With the goal of reducing unnecessary health care expenditures, some health insurance companies in Germany offer plans with conditional refunds as monetary incentives for the insured. Specifically, contingent on being claim-free for a full calendar year, a client receives a certain multiple of monthly premiums back depending on the health care plan and the number of previous claim-free years. Conceptually, the premium refunds can be seen as a deferred reward from non-utilization rather than a instantaneous punishment from utilization, but otherwise having the same function as a deductible. These dynamic incentives hence offer an alternative way of studying the behavioral responses from dynamic incentives in non-linear insurance contracts.[1]

[1] Zweifel (1987, 1992) have previously studied the impact of premium refunds in the German private health insurance system and found that such incentives are indeed associated with reduced health care costs. How-

We use unique claims data from a large German private health insurer to empirically investigate how the option of receiving a conditional premium refund affects individual claiming behavior in both the short and the long run. To claim causality we exploit an insurer policy that significantly increased the refunded amount in some insurance plans while leaving other plans unchanged, implemented by comparing individuals in insurance plans affected and unaffected by the insurer policy in a difference-in-differences framework. Furthermore, we suggest an intuitive way to decompose the overall impact of the insurance policy into an intensive, an extensive and an automatic effect relating to the total amount claimed, the propensity of claiming and the mechanical impact of raising the co-payment threshold, respectively. Finally, we also provide a number of extensions to our main analysis in order to study the mechanisms and robustness of our results.

Our findings show that the average annual claim by a client is significantly reduced by €240 (8%) as a consequence of the increased refund insurer policy. Our decomposition approach reveals that claims are reduced in all three components and less than one-third of the total amount is explained by the 'automatic' effect. Thus, we find evidence that clients react to the changed incentives from the increased refund size by changing their claiming behavior. Furthermore, when splitting the result into subgroups of sex, age, risk and occupation we find that the main result is robust for all groups and that women, older, high-risk and self-employed clients appear particularly responsive to the changed incentives. We also find that the insurer incurred a loss from the policy by analyzing the revenues and costs from changed reimbursements and refund payouts. Finally, studying the long term impact of the insurer policy we find some evidence of increased health expenditures in later years, suggesting that individuals did not only reduce unnecessary spending.

Our findings contribute to the research on client behavior in non-linear insurance contracts in several ways: First, we find evidence that individuals respond to changes in future prices also when the monetary incentive is framed as a deferred reward rather than an instantaneous punishment. Second, our decomposition approach reveals that the effect above the new threshold is mainly driven by the extensive margin suggesting that price-shopping

ever, these findings may to some extent be driven by selection of healthier individuals into contracts with refund options which cannot be ruled out completely in the author's data.

may be limited and deferred utilization more important. This is also indicated by our finding that expenditures appear to increase in the long run. The results from our study add to the previous evidence showing that individuals appear to understand complex dynamic incentives which are likely to result in non-trivial aggregate effects on costs of health care. However, while policymakers and insurance providers should take such behavior into account when designing effective insurance policies, they should also bear in mind that such policies may not only decrease unnecessary expenditures.

The remainder of this paper is organized as follows. In the next section we describe the institutional setting in greater detail, while the empirical approach is laid out in Section 3. Section 4 introduces the data, while Section 5 contains our results and Section 6 concludes.

2 Institutional Setting and Research Design

2.1 The German Health Care System

In the German health system a statutory and a private health insurance market coexist.[2] For the majority of the German population statutory health insurance is mandatory while certain groups may opt out for private health insurance. This exemption applies to civil servants, self-employed, freelancers and high-income earners.[3] As of 2014, for example, roughly 11% of the German population held a full health insurance contract with a private insurance company (PKV, 2015).

One notable difference between the two systems is the calculation of the monthly premium. The premium in the statutory health insurance is solely based on the earnings of an individual and is by federal law not allowed to vary across the statutory insurance funds. The total premium amounts to roughly 15.5% of gross salary of which half is paid by the employer as social contributions (together with pension and unemployment contributions). In contrast, individual income is disregarded in the premium calculations in the private health insurance system. Instead, premiums are risk-adjusted according to some predefined

[2]For a comparison of the two systems see, e.g., Karlsson *et al.* (2016).

[3]The annual income threshold *(Versicherungspflichtgrenze)* is adjusted on a yearly basis. In 2014 it was equal to € 53,550.

schedule specific to the different insurance firms for which they are free to compete over to attract customers. Risk factors such as age, gender and health risk determine the individually calculated premium. Furthermore, most private insurance contracts contain a bundle of services, coverage, and reimbursement rates from which the client can flexibly include or exclude in order to adjust the premium and benefit package in accordance with his or her needs. The level of the deductible is one such important choice variable. In contrast, contracts in the statutory health insurance include an, in all relevant aspects, uniform benefit package and contain no forms of co-payments.

2.2 Premium Refunds

A unique aspect of the private health insurance system in Germany is that many companies offer premium refunds. Contingent on being claim-free for a whole calendar year, an insured individual may receive a multiple of monthly premiums back from the insurer. Such refunds are often designed in a dynamic fashion. The number of monthly premiums that might potentially be gained increases with number of previous claim-free years. An insurance plan client immediately disqualifies for such a refund once submitting the first claim for a particular calendar year. This even applies when the respective claim is not reimbursed, for example due to the fact that its amount is lower than the respective deductible. On the contrary, if there was not a single claim for a particular calendar year, the insurer automatically transfers the refund to the client at some predetermined date in following year.

Broadly speaking, there exist two different types of refunds. The first one is dependent on the economic success of the respective insurance company. These refunds are not explicitly specified in the respective health insurance policies, but the insurer decides on a voluntary basis from year to year to pay out these refunds. The majority of refunds in the private health insurance sector are of that first type. In 2014, for example, they made up for about 92 percent of all refunds (PKV, 2015, p. 44). The refunds of the second type are different. In contrast to the first type they are explicitly written into the respective policies and their payout is therefore guaranteed and independent of the business success of the insurer. We study data of a large private health insurer in Germany in this paper. For this particular insurer the relative importance of the two refund types is similar to overall figure of the sector as a whole. However, in this paper we only analyse the first type of refunds.

There exists heterogeneity in the sector of private health insurers on how to treat different types of health care services with respect to refunds. Sometimes, specific services might be disregarded, such as inpatient or preventive care. That means, clients are reimbursed for those specific services but nevertheless qualify for premium refunds still in such settings. However, such exemptions are not present for any of the policies that we study in this paper.

2.3 Research Design

Table 1 provides an overview of our research design. The intervention we study is a sudden increase in premium refunds in certain plans. The change was announced in February 2008 and applied to claims made during the calendar year of 2008.

TABLE 1.
Sample overview

Group	Plans	Refund size				Observations
		2005-2007	2008	2009	2010-2011	
Treatment	6	0.5/1/1.5/2	3	1/3	1/2/3	46,359 (55%)
Treatment	2	0.5/1/1.5/2	3	1/3	–	1,755 (2%)
Control	6	0.5/1	0.5/1	–	–	35,572 (43%)

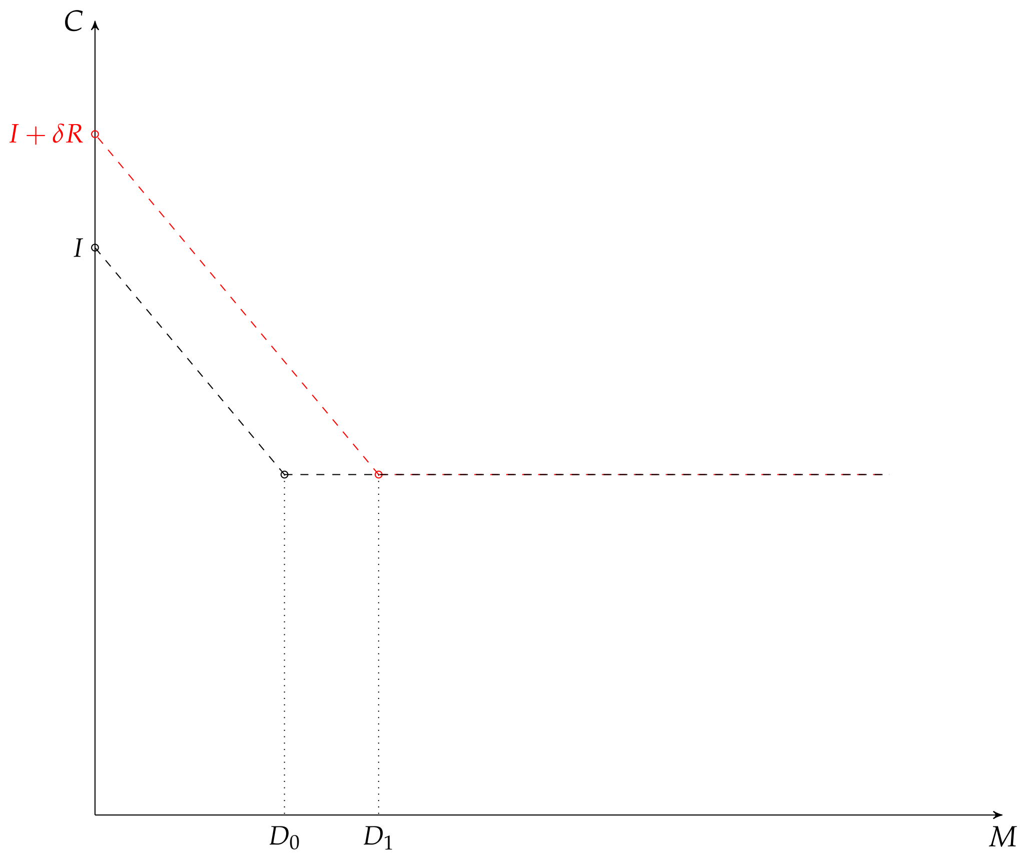
NOTE.— Refund size measured in monthly premiums if claim-free for one/two/three/four consecutive years. Listed by year until maximum is reached.

Our treatment group consists of more than 48,000 clients from 8 specific plans for which the insurer changed the refund structure in 2008. Before the refund policy changed in 2008, an experience-rated refund was offered to all plans. An additional year that a client refrains from filing a claim is rewarded with an additional refund of 0.5 monthly premiums, up to a maximum of 2 premiums. Accordingly, for instance, a client who does not file any claim for 3 consecutive years receives a refund of 1.5 monthly premiums. In 2008 the refund structure for these plans changed and every client instead received a refund of 3 monthly premiums if no claims were filed, regardless of the previous claiming history. As our control group we select more than 35,000 clients from 6 different plans with a constant refund structure from 2005 to 2008. Throughout the study period these clients are subjected to an experience-rated refund scheme that rewards the first year without claims by half a monthly premium and every following year with one full monthly premium.

2.4 Theoretical Predictions

In Figure 2, we present a simple sketch of how the sudden change in the refund scheme affected the budget set available to clients. The Figure shows possible combinations of consumption of health care (M) and any other commodities (C), before and after the intervention (for simplicity we have assumed that there was no refund scheme to begin with and normalised all prices to one).

FIGURE 2.
Budget Sets before and after Refund Increase



Prior to the intervention, the client pays the full cost of care up to the level of the deductible D_0, after which health care costs are fully reimbursed. The intervention is equivalent to an increase in the client's income, where R is the amount of refund they can expect after a claim-free year, and δ the individual's time discount factor – reflecting the fact that the premium refund is paid out several months after the end of the calendar year. Clearly, $\delta R = D_1 - D_0$. In terms of budget constraints, the refund policy has an impact very similar to that of the deductible D_0 – but it is not necessarily associated with a reduction in premiums, as the cost savings are balanced by the refund payments.[4]

As mentioned above, the rationality of consumers in the sequential decision problem that health care utilisation represents is a hotly debated issue in the literature, in particular when there are dynamic incentives to consider as in the case of these premium refunds. Nevertheless, it is useful to consider the expected response of a rational consumer with well-behaved preferences in a static framework as a benchmark. We thus now introduce a simple model similar to that of Dalton (2014), with a utility function $U(M, C \mid H)$ which determines the utility associated with different care and consumption bundles (M, C) depending on their health status H. We assume that the function is concave whenever H is below perfect health – and when $H = H^{max}$ the marginal utility with respect to health care consumption may be equal to 0. We also assume that whenever the individual is not in perfect health ($H < H^{max}$), their marginal rate of substitution is larger than the price of health care services when the amount consumed approaches zero:

$$\forall\, C > 0\ \exists\, \epsilon > 0 : MRS(M, C) = \frac{\partial U(M, C \mid H) / \partial M}{\partial U(M, C \mid H) / \partial C} > 1 \text{ if } M < \epsilon \text{ and } H < H^{max} \quad (1)$$

Based on these basic assumptions, it is possible to produce some very useful comparative statics regarding how consumers can be expected to respond to the change in refunds. One prediction follows immediately from an inspection of the budget constraint in Figure 2: no consumer whose optimal bundle is in the range $M \in [0, D_1)$ without the reform will find it optimal to consume at $M > D_1$ after the reform. This is a direct consequence of the strong axiom of revealed preference: all bundles with $M > D_1$ were available before the reform and the consumer preferred a different bundle which is still feasible.

[4]An additional difference between premium refunds and deductibles has been highlighted by Zweifel (1992): a refund scheme offers the client some degree of 'consumption smoothing' – or rather: *utility* smoothing – as they have the opportunity to temporally dissociate the financial risk from the health risk.

In addition, our assumption in equation (1) above regarding the MRS as M approaches zero allows us to rule out another possibility: no consumer with health care utilisation $M \in (0, \infty)$ will find it optimal to switch to $M = 0$ in the presence of premium refunds. Conversely, no consumer with $M = 0$ to begin with will find it optimal to change their health care utilisation either, since they are assumed to be in perfect health and thus not deriving utility from health care consumption. Consumers with utilisation $M \in (0, D_0)$ under the old regime will *increase* their health care consumption: this is a pure income effect from the increase in premium refunds. Consumers with utilisation in the region $M \in (D_0, D_1)$ may move to any other point in the segment (D_0, D_1): they also experience an increase in their incomes, but also an increase in the relative price of health care. Finally, consumers who are initially in the segment $M > D_1$ will stay were they are in case they are in relatively poor health so that the utility loss of downsizing their health care utilitsation to below D_1 is large. On the other hand, some consumers in relatively good health (and thus consumption not far above D_1) might find it rational to shift consumption to the segment $(0, D_1)$.

Thus, based on some very basic and plausible assumptions regarding the preferences of consumers, we may state a number of hypotheses regarding the reactions to a change in premium refunds:

1. The probability of observing claims $M > D_1$ should decrease.

2. The intensity of claims above the new threshold D_1 should increase.

3. The proportion of positive claims (as opposed to zero claims) below the threshold will increase.

4. The intensity of positive claims below D_1 is likely to increase.

As will be outlined below, hypotheses (1) and (2) relate to directly observable facts. Hypothesis (3) and (4) relate to things that cannot be directly observed in the data – and in addition, hypothesis (4) does not follow automatically from the assumptions, since a reduction is conceivable even if unlikely. Nevertheless, hypothesis (3) and (4) are useful in order to make predictions of how the entire distribution of health care utilisation will adjust. In particular if we find empirical evidence consistent with hypotheses (1) and (2), we can use hypotheses (3) and (4) to conduct preliminary inference regarding how unobserved utilisation below D_1 changes.

These predictions may fail to materialise for a number of reasons. We have already mentioned the case of complete myopia. But even if the client is rational and perfectly forward-looking, they may not have the degree of control over their health care utilisation as the theory assumes. It is common to assume that the consumer decides on the initiation of a health care episode, but has limited autonomy in subsequent decisions (cf. Ellis and Zhu, 2016; Aron-Dine *et al.*, 2013; Keeler and Rolph, 1988). Finally, the consumer might be so sophisticated that they optimise the timing of their health care utilisation. For example, having utilisation $M' > 2D_1$ in one year and zero in the next year, is clearly preferable over having utilisation equal to $M'/2$ in both years – as long as the bunching of utilisation in one year does not have too severe consequences for the individual's health (Cabral, 2016).

3 Empirical Approach

3.1 Definitions

The variable of primary interest in our analysis is total annual health care expenditure. Denote by M_{it} the annual expenditure of an arbitrary individual i in year t. The distribution of M_{it} within a group of clients holding a certain contract will typically be of mixed type with a mass point at zero. We denote the cdf of such a distribution by $F(M)$. The intervention we are considering is an increase in the rebate for zero claims. For an evaluation of the intervention, it would be of interest to compare how the distribution $F(M)$ – or any statistic based on it, such as e.g. the mean utilisation – changed in response to the reform. Defining the counterfactual distribution in the absence of treatment as $F^0(M)$ we could, for example, define the average treatment effect on the treated (ATT) as:

$$ATT = \int_{m=0}^{\infty} m \left[\mathrm{d}F(m) - \mathrm{d}F^0(m) \right] \tag{2}$$

A main challenge in our analysis is that the dataset is a claims dataset and therefore M_{it} is only imperfectly observed. The annual claims Y, which are observed, are typically a known function of actual utilisation: $Y(M_i)$. However, claims are censored at a certain level whenever the contract entails a deductible and/or a rebate. This censoring has a number of consequences. First, the averages forming the components of (2) will never be observed in our data since a part of the distribution will be missing. Second, the censoring removes comparability between observations that have been exposed to different contract parameters. This represents an additional problem since the intervention we consider is a change in such contract parameters. We will however show below that it is nevertheless possible to draw meaningful inference based parts of the distribution of M that are unaffected by the censoring.

We denote by D_{it} the sum of the deductible applying in a certain contract, and the rebate applying after a claim-free year.[5] It is the sum D_{it} which determines whether it is rational to submit claims for positive health care expenditure at the end of the year. Consider for example an individual who has a deductible of €500 and a potential rebate of €300. If health care expenditure amounts to €700 in a certain year, they will get €200 reimbursed if they submit their bills – and instead a rebate of €300 if they do not. For total health care expenditure of €900 it would on the other hand be rational to submit the bills; at least in the absence of dynamic incentives. Thus, we may write the claiming function as:

$$Y(M) = \begin{cases} 0 & \text{if } M \leq D \\ M & \text{if } M > D. \end{cases} \tag{3}$$

The intervention we consider is a sudden significant increase in D_{it} that occurred in 2008: the deductible part of D_{it} remained constant whereas the rebate increased by one or more

[5] D_{it} is partly based on claims made in previous years, and on an individual's premium, and thus it varies at the individual level. In what follows, we will nevertheless suppress the individual-level variation: the exposition may be thought of as implicitly conditioning on personal characteristics such as the individual premium.

monthly premiums. From the point of view of the insurance company, we might define the causal effect of this intervention as

$$ATT^{y} = \int_{m=0}^{\infty} Y^{1}(m)\, \mathrm{d}F(m) - \int_{m=0}^{\infty} Y^{0}(m)\, \mathrm{d}F^{0}(m) \tag{4}$$

where we have given ATT the superscript y to highlight that it applies to claims only (and the function $Y(M)$ is given different superscripts depending on treatment status). Even though the effect presented in equation (4) undoubtedly represents a causal effect – namely by how much the intervention would change the total claims made by clients in a certain contract – it is of limited interest from an economic point of view. The effect ATT^{y} combines two different responses to the intervention in an obscure way: an automatic effect arising because clients now claim according to function $Y^{1}(m)$ due to the higher rebate applying, and a behavioural (or moral hazard) effect arising because the new incentives have moved the distribution of utilisation from $F^{0}(m)$ to $F(m)$.

However, some rearrangement of equation (4) actually allows us to disentangle the two effects:

$$\begin{aligned}
ATT^{y} &= \int_{m=0}^{\infty} Y^{1}(m)\, \mathrm{d}F(m) - \int_{m=0}^{\infty} Y^{0}(m)\, \mathrm{d}F^{0}(m) \\
&= \int_{m=D_1}^{\infty} m\mathrm{d}F(m) - \int_{m=D_0}^{\infty} m\mathrm{d}F^{0}(m) \\
&= \int_{m=D_1}^{\infty} m\mathrm{d}F(m) - \int_{m=D_1}^{\infty} m\mathrm{d}F^{0}(m) - \int_{m=D_0}^{D_1} m\mathrm{d}F^{0}(m) \\
&= \left(1 - F(D_1)\right) \mathbb{E}_1\left[M \mid M > D_1\right] - \left(1 - F^{0}(D_1)\right) \mathbb{E}_0\left[M \mid M > D_1\right] - \\
&\quad \int_{m=D_0}^{D1} m\mathrm{d}F^{0}(m) \\
&= \left(1 - F^{0}(D_1)\right) \left[\mathbb{E}_1\left[M \mid M > D_1\right] - \mathbb{E}_0\left[M \mid M > D_1\right]\right] && \text{(INT)} \\
&\quad + \left(F^{0}(D_1) - F(D_1)\right) \mathbb{E}_1\left[M \mid M > D_1\right] && \text{(EXT)} \\
&\quad - \int_{m=D_0}^{D_1} m\mathrm{d}F^{0}(m) && \text{(AUT)}
\end{aligned}$$

where $\mathbb{E}_1$ ($\mathbb{E}_0$) represents expectations taken in the presence (absence) of treatment. We have split up the overall effect on claims, ATT^{y}, into three parts: an *intensive* margin, capturing by how much claims above the new rebate and deductible level change on average, an *extensive* margin, representing the probability of reaching the new level D_1, and an *automatic* component, which simply reflects that total expenditures between D_0 and D_1 will not be claimed anymore.[6] The distinction between automatic effects and utilisation effects is necessary in order to say anything about moral hazard. The distinction between the intensive and the extensive margin is also very useful since it corresponds directly to hypotheses (1) and (2) presented in section 2.4 above.

3.2 Identification

Splitting the total effect on claims into automatic and behavioural changes seems to solve some of the problems related to the censoring in claims. However, we also need to deal with the problem that one component in the definition of ATT^y, the counterfactual $\int_{m=0}^{\infty} Y^0(m)\,\mathrm{d}F^0(m)$ is unobserved, since it represents the claims that had been submitted if the change in rebates had never taken place. In order to solve the identification problem, we use a difference-in-difference design and use a control group to impute the missing counterfactual. We will thus use as our control group contracts that were not exposed to an increase in rebates in 2008.

Since it is our aim to decompose the overall effect of the intervention on claims into three parts, the conditions for identification are also stronger than in a standard DiD setting. For any statistic $G(\cdot)$ based on the counterfactual distribution $F^0(m)$ used in the decomposition of the treatment effect above, we build a counterfactual based on trends in the control group:

$$G\left(F_{11}^0(m)\right) = G(F_{10}(m)) + [G(F_{01}(m)) - G(F_{00}(m))] \tag{5}$$

where we have added subscripts to $F(m)$ representing groups and periods; for example, $F_{10}(m)$ is the distribution of expenditures in the treatment group before the intervention – a distribution which is observed whenever $M > D_0$.

[6] It should be noted that the wording is slightly misleading since the 'automatic' effect may also arise due to changes in utilisation. It does, however, seem reasonable to assume that most of it is driven by changes in claiming behaviour.

Identification of effects is thus based on three component-specific common time trend assumptions and one overall common time trend assumption. The latter may be defined as:

$$\int_{m=D_0}^{\infty} m\mathrm{d}F_{11}^{0}\left(m\right) - \int_{m=D_0}^{\infty} m\mathrm{d}F_{10}\left(m\right) = \int_{m=D_c}^{\infty} m\mathrm{d}F_{01}\left(m\right) - \int_{m=D_c}^{\infty} m\mathrm{d}F_{00}\left(m\right) \tag{CT}$$

where D_c denotes the sum of deductible and rebate applying in the control group (and which does not change over time). It should be obvious from equation (CT) that the assumption is more likely to be satisfied if D_c is of similar size as D_0; however, this condition is neither necessary nor sufficient for the (CT) assumption to hold.

The overall common time trend condition is mirrored by three component-specific conditions which need not be satisfied even when the overall condition (CT) is satisfied. For the intensive margin effect (INT), the condition for identification would be:

$$\int_{m=D_1}^{\infty} m\mathrm{d}F_{11}^{0}\left(m\right) - \int_{m=D_1}^{\infty} m\mathrm{d}F_{10}\left(m\right) = \int_{m=D_1}^{\infty} m\mathrm{d}F_{01}\left(m\right) - \int_{m=D_1}^{\infty} m\mathrm{d}F_{00}\left(m\right) \tag{CT-INT}$$

which simply states that the expected value of expenditure above the *new* level D_1 would follow a common time trend in the absence of the intervention. It is clear from (CT-INT) that identification requires $D_1 > D_c$ since otherwise the right-hand side of the equation would be unobserved.

The two remaining common trend assumptions may finally be stated as

$$F_{11}^{0}\left(D_1\right) - F_{10}\left(D_1\right) = F_{01}\left(D_1\right) - F_{00}\left(D_1\right) \tag{CT-EXT}$$

for the extensive margin effect (EXT) and as

$$\int_{m=D_0}^{D_1} m\mathrm{d}F_{11}^{0}\left(m\right) - \int_{m=D_0}^{D_1} m\mathrm{d}F_{10}\left(m\right) = \int_{m=D_c}^{D_1} m\mathrm{d}F_{01}\left(m\right) - \int_{m=D_c}^{D_1} m\mathrm{d}F_{00}\left(m\right) \tag{CT-AUT}$$

for the 'automatic' effect (AUT). Again, it is notable that assumption (CT-AUT) is more likely to hold whenever $D_c \approx D_0$ – whereas the two other component-wise time trend assumptions are likely to be less sensitive to deviations of D_c from D_0.

We also compare the results from our difference-in-differences analysis to alternative assumptions on the underlying time trends of the utilization in our data. Specifically, we relax the three common trends assumptions listed above in favor of a simple linear before-after trend defined in three different ways: a common linear trend for all individuals, a plan specific time trend and an individual-specific trend. In most of what follows we show the results for both the difference-in-differences approach and the before-after model.

3.3 Estimation

The overall effect of the intervention on total annual claims – ATT^y defined in equation (4) – may be estimated using standard difference-in-differences. Thus, in our first specification, we regress total annual claims Y_{it} on year dummies, a dummy variable for the treatment group, and on an interaction term representing individuals in the treatment group in the post-treatment period. Furthermore we control for age (linear and quadratic term) as well as for the presence of a risk premium and female sex. We also distinguish self-employed and white-collar workers from other types of occupation. Under the assumption stated above, this regression would identify ATT^y or the total effect of the policy on annual claims.

In order to identify *behavioural* changes, we generate the new variable Y'_{it}, which is censored at the new premium and deductible level D_1 according to the following formula:

$$Y'_{it} = \begin{cases} 0 & \text{if } Y_{it} \leq D_1 \\ Y_{it} & \text{if } Y_{it} > D_1. \end{cases} \tag{6}$$

Using this new outcome variable in the DiD regression characterised above identifies the combined behavioural response (EXT)+(INT). It follows that the difference between estimates from these two specifications identifies the automatic effect (AUT).

In order to finally partition the behavioural response above D_1 along the intensive and extensive margins, we define a third variable Y''_{it} as follows:

$$Y''_{it} = \begin{cases} 0 & \text{if } Y_{it} \leq D_1 \\ 1 & \text{if } Y_{it} > D_1. \end{cases} \tag{7}$$

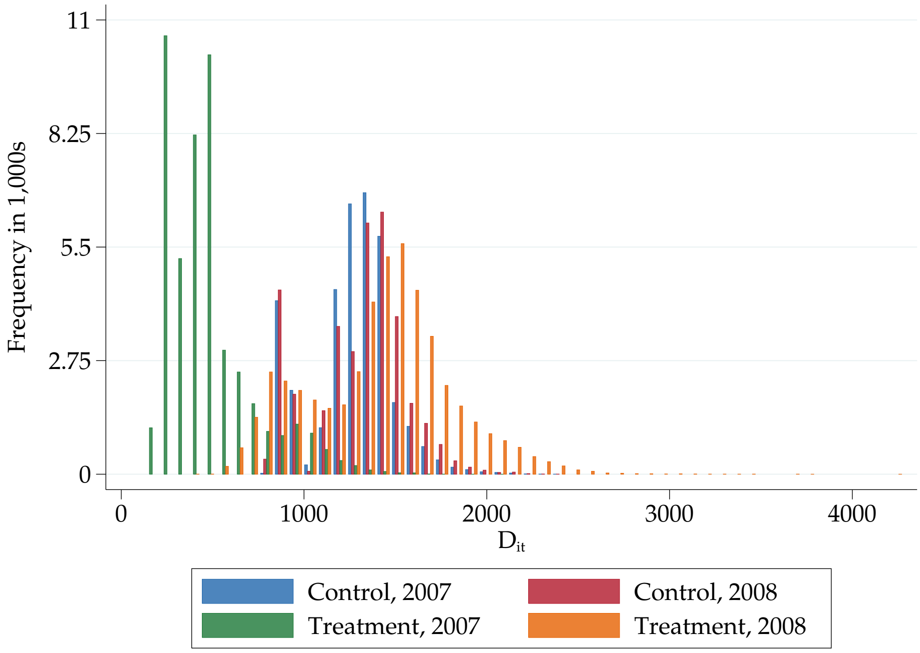
Using Y''_{it} as an outcome variable and multiplying the DiD estimate with $\mathbb{E}_1 [M \mid M > D_1]$ – i.e. the expected total costs above the new rebate and deductible level D_1 after the intervention – would give us the *extensive* margin effect (EXT). Finally, subtracting this estimate from our previous estimate of (EXT)+(INT) would also give us the *intensive* margin effect (INT).

Since the distribution of the error term is unknown and some estimates are based on calculating differences between estimates, we use bootstrapping with 400 replications for statistical inference.

4 Data

We exploit rich claim-level data from a large private health insurer in Germany from the years 2005 to 2008 in our main analyses. In order to analyze the effects of the refund structure we collapse information to the person-year-level. Furthermore, we restrict our data in various ways. Firstly, we only consider clients aged 30 or older with non-missing information in order to rule out additional complications arising with insured children and families. Secondly, the sample is restricted to clients that hold their contract throughout the study period from 2005 to 2008, i.e. switching between plans as well as new entries are not allowed. This implies that we work with a balanced panel of clients. 2008 is the year in which premium refunds are increased. We do not consider later years as these bring about additional changes to the design of the health insurance plans. However, in a subsequent analysis we study the long-run impact of the refund policy for a specific subsample in which the refund were abolished in 2010. All monetary variables (premiums, deductibles, claims, benefits, refunds) have been adjusted to account for inflation and expressed in 2011 euros using the CPI provided by the German Federal Statistical Office.

FIGURE 3.
Distribution of D_{it}



NOTE.— Claims data, 2005 – 2008.

One crucial variable for our analysis is the individual-specific threshold D_{it} calculated as the sum of the deductible and the conditional refund. Figure 3 illustrates how D_{it} is distributed in our sample for 2007 and 2008 by treatment status. The increase of refunds in the treatment group between 2007 and 2008 is visible while the distributions of the threshold values in the control group are very similar across years.[7]

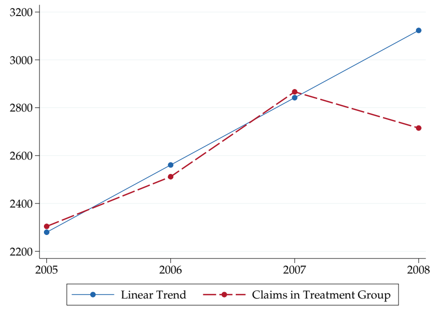
Figure 4 provides an overview on the evolution of our three claim component outcomes – overall claims (Y), amount claimed above the new threshold (Y') and probability of claiming above the new threshold (Y'') – by group and empirical approach. Specifically, the left figures show comparisons between the trends in the treatment group and a common linear trend while the right figures compare the treatment and control groups for the DiD analysis. The simple linear trend is able to fit the pre-trend in the treatment group surprisingly well

[7]Deductibles are gradually increased by the insurer over time in order to account for inflation. One might consider to restrict the sample to those plans with constant deductibles in 2005–2008 only. We do so in a sensitivity analysis with qualitatively similar results.

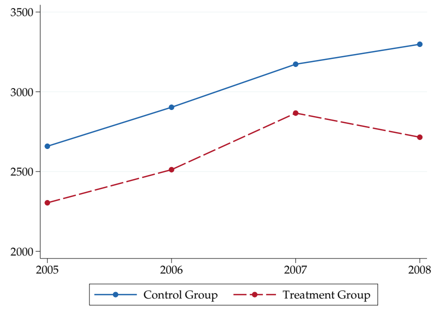
for all three components. Furthermore, the common trend assumptions in the DiD specification also seem to be valid and formal tests of the pre-trends confirm this observation. The figure also shows clear evidence that there is a clear trend break in 2008 for the treatment group, suggesting that the new refund policy had an impact on the claiming behavior of clients.

FIGURE 4.
Graphical evidence of pre-trends for claim components

Mean claims, $\mathbb{E}[Y]$

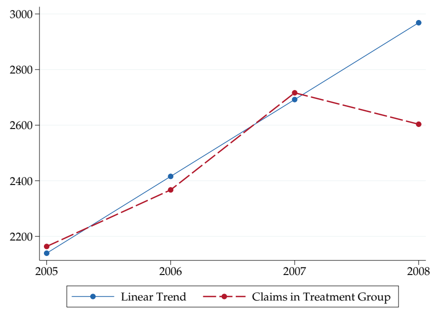


(a)
Before-After

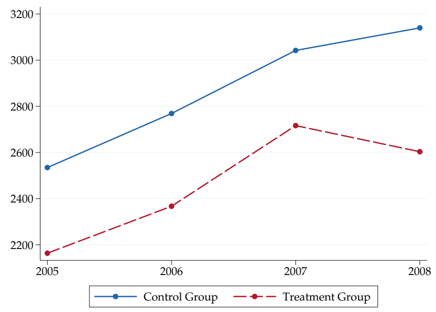


(b)
DiD

Mean of claims censored at D_1, $\mathbb{E}[Y']$

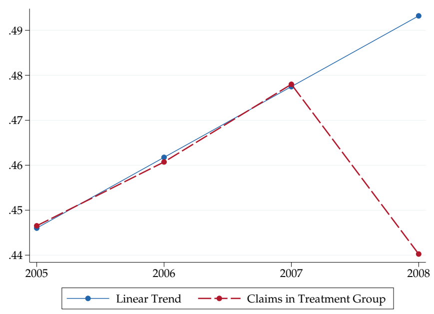


(a)
Before-After

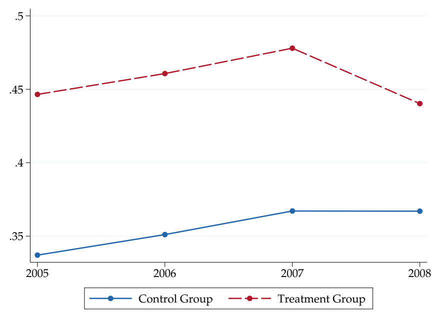


(b)
DiD

Average probability of claiming above D_1, $\mathbb{E}[Y'']$



(a)
Before-After



(b)
DiD

NOTE.— Claims data, 2005 – 2011. D_1 = new refund level (i.e. premium + deductible).

Table 2 reports summary statistic of the remaining variables in our sample for 2008. Insurees in the control group are more likely to be female, older, to have a higher deductible but a lower insurance premium. Furthermore, the control group is mostly represented by self-employed while the majority of clients in the treatment group are white-collar workers. One may be concerned about the significant differences across the treatment and control group with respect to their claiming behavior. However, as long as these differences only manifest in differing outcome levels our difference-in-difference approach will nevertheless produce valid estimates.[8]

TABLE 2.
Descriptive statistics, 2008

	Treated		**Controls**	
	mean	sd	mean	sd
Female	0.11	0.31	0.30	0.46
Age	44.08	7.68	48.90	10.14
Premium	409.04	150.65	373.55	179.57
Deductible	182.68	147.22	1053.97	151.42
Risk premium	0.32	0.47	0.32	0.47
Self employed	0.32	0.47	0.69	0.46
White-collar	0.65	0.48	0.23	0.42
Other occupation	0.03	0.18	0.08	0.27
Observations	48,114		35,572	

NOTE.— Claims data, 2005 – 2008. All monetary variables measured in 2011 euros.

5 Results

5.1 Main Results

Table 3 presents our main estimates of the total effects of the intervention and its decomposition. Overall, as the total effect in column (1) indicates, the increase of refunds significantly reduces the average amount claimed by about €240–420 depending on specification. In the absence of treatment, the average amount claimed in 2008 would have been around €3120 (€2960) in the treatment group which corresponds to a reduction of about 13% (8%) at base-

[8]Table A.1–Table A.3 provide summary statistics for the years 2005–2007.

line for the before-after (DiD) model. The total effect is the sum of the component-specific effects in columns (2)–(4) in which all three indicate reduced claiming as a consequence of the new refund policy. First, column (2) reports the results for the intensive margin. Here, the model-specific results are somewhat different with a substantially higher, and significant, effect €235 for the before-after model compared to an insignificant €85 in the DiD model. In contrast, the reduction at the extensive margin in column (3) is significant and similar in magnitude for both specifications amounting to about one third of the overall reduction in claims. Finally, the automatic effect in column (4) is slightly higher for the DiD model but account for a relatively small share of the total effect in both specifications.

TABLE 3.
Decomposition results by model specification

	Total	=	Intensive Margin	+	Extensive Margin	+	Automatic
	(1)		(2)		(3)		(4)
Panel A. Before-after model ($n = 48,114$)							
Effect	-417.58		-235.70		-139.11		-42.78
(s.e.)	(65.14)		(64.67)		(8.35)		(2.64)
%	100%		56%		33%		10%
Baseline				3123.15			
% of Baseline	-13.37%		-7.55%		-4.45%		-1.37%
Panel B. DiD model ($n = 83,686$)							
Effect	-244.10		-84.78		-93.40		-65.93
(s.e.)	(68.48)		(67.13)		(9.60)		(3.90)
%	100%		35%		38%		27%
Baseline				2959.30			
% of Baseline	-8.25%		-2.86%		-3.16%		-2.23%

NOTE.— Claims data, 2005-2008. Estimates obtained after combined approach of three before-after-regressions (panel A) or three difference-in-difference regressions (panel B), see text for details. Standard errors in parentheses, obtained after bootstrapping with 400 replications. Baseline value refers to the average claimed amount that would have occurred in the treatment group in the treatment period (2008) in the absence of treatment. Control variables in all regressions: age (linear and quadratic term), dummies for the presence of a risk premium, female gender, self-employment or white-collar workers, distinguishing plans that are only for physicians from other plans.

Table 4 summarizes the corresponding decomposition results by specification. The first three columns correspond to a before-after-comparison with either a common linear, plan-specific or individual-specific time trends. Column (4) reproduces the DiD results for comparison. The results are largely robust to the choice of trends in the before-after analysis.

Taken together, the results suggest a significant overall effect that seems to be mainly driven by reductions at the extensive margin and according to the automatic effect in the DiD model. Furthermore, the before-after-comparison seems to some extent overstate the size of the effects compared to the more conservative DiD results. In the following we will therefore focus on the results derived from the DiD model.

TABLE 4.
Summary of decomposition results for different specifications

	Before-After: Trends ($n = 48,114$)			**DiD** ($n = 83,686$)
	Common (1)	Plan (2)	Individual (3)	(4)
Total	-417.58 (65.14)	-417.89 (65.13)	-430.29 (70.63)	-244.10 (68.48)
Intensive Margin	-235.70 (64.67)	-235.90 (64.66)	-248.26 (70.36)	-84.78 (67.13)
Extensive Margin	-139.11 (8.35)	-139.24 (8.35)	-138.79 (8.51)	-93.40 (9.60)
Automatic	-42.78 (2.64)	-42.75 (2.65)	-43.24 (2.71)	-65.93 (3.90)

NOTE.— Claims data, 2005-2008. Estimates obtained after combined approach of three before-after-regressions (columns 1–3) or three difference-in-difference regressions (column 4), see text for details. Standard errors in parentheses, obtained after bootstrapping with 400 replications. Baseline value refers to the average claimed amount that would have occurred in the treatment group in the treatment period (2008) in the absence of treatment. Control variables in all regressions: age (linear and quadratic term), dummies for the presence of a risk premium, female gender, self-employment or white-collar workers, distinguishing plans that are only for physicians from other plans.

5.2 Heterogeneity by subgroups

In this section we study whether certain subgroups (sex, age, risk group and occupation) are particularly sensitive to the change in incentives brought about by the intervention. Table 5 presents the results by sex. The table suggest that females claim more on average, respond more strongly to the intervention than males, with significant and sizable extensive and intensive margin effects. Since females also claim more on average the difference in percentages from the baseline are somewhat smaller but the overall pattern remains.

TABLE 5.
Decomposition results by sex: DiD model

	Total	=	Intensive Margin	+	Extensive Margin	+	Automatic
	(1)		(2)		(3)		(4)
Panel A. Males ($n = 67,634$)							
Effect	-226.70		-74.32		-83.30		-69.07
(s.e.)	(72.91)		(71.20)		(10.58)		(4.68)
%	100%		33%		37%		30%
Baseline				2,800			
% of Baseline	-8.09%		-2.65%		-2.97%		-2.47%
Panel B. Females ($n = 16,052$)							
Effect	-440.15		-186.94		-191.73		-61.47
(s.e.)	(117.37)		(109.72)		(32.83)		(10.64)
%	100%		42%		44%		14%
Baseline				4,296			
% of Baseline	-10.24%		-4.35%		-4.46%		-1.43%

NOTE.— Claims data, 2005-2008. Estimates obtained after combined approach of three difference-in-difference regressions, see text for details. Panels refer to different subgroups. Standard errors in parentheses, obtained after bootstrapping with 400 replications. *Baseline* value refers to the average claimed amount that would have occurred in the treatment group in the treatment period (2008) in the absence of treatment. Control variables in all regressions: age (linear and quadratic term), dummies for the presence of a risk premium, female gender, self-employment or white-collar workers, distinguishing plans that are only for physicians from other plans.

Table 6 and Table 7 present analogue results by age (above and below median age, respectively) and by risk category and occupation (having a risk premium or not, being self-employed or white-collar worker, respectively). The results by age suggests that older individuals responded more strongly in both relative and absolute terms than younger individuals. However, the variation is also much higher for older clients rendering the effect on the intensive margin insignificant. Furthermore, individuals with a risk premium respond stronger than individuals with no risk premium and self-employed respond much stronger than white-collar workers. Interestingly, self-employed have the highest percentage effect of all subgroups, reducing their claims by an average of more then 15%. This may be related to the fact that they are likely to have more insight in the system due to that they have to manage their insurance by themselves compared to other occupational groups. In summary, for all subgroups we find significant and sizable total effects, extensive margin effects as well as automatic effects. The general pattern for the full sample is largely robust also within subgroups.

TABLE 6.
Decomposition results by age: DiD model

	Total	=	Intensive Margin	+	Extensive Margin	+	Automatic
	(1)		(2)		(3)		(4)
Panel A. Age in 2005 ≤ 41 ($n = 42,523$)							
Effect	-173.56		-31.67		-85.52		-56.36
(s.e.)	(50.96)		(47.98)		(10.71)		(4.88)
%	100%		18%		49%		32%
Baseline				2,322			
% of Baseline	-7.47%		-1.36%		-3.68%		-2.43%
Panel B. Age in 2005 > 41 ($n = 41,163$)							
Effect	-336.33		-144.85		-118.11		-73.36
(s.e.)	(136.36)		(125.38)		(22.76)		(6.30)
%	100%		43%		35%		22%
Baseline				3,879			
% of Baseline	-8.67%		-3.73%		-3.04%		-1.89%

NOTE.— Claims data, 2005-2008. Estimates obtained after combined approach of three difference-in-difference regressions, see text for details. Panels refer to different subgroups. Standard errors in parentheses, obtained after bootstrapping with 400 replications. *Baseline* value refers to the average claimed amount that would have occurred in the treatment group in the treatment period (2008) in the absence of treatment. Control variables in all regressions: age (linear and quadratic term), dummies for the presence of a risk premium, female gender, self-employment or white-collar workers, distinguishing plans that are only for physicians from other plans.

TABLE 7.
Decomposition results by risk premium and occupation: DiD model

	Total	=	Intensive Margin	+	Extensive Margin	+	Automatic
	(1)		(2)		(3)		(4)
Panel A. No Risk Premium ($n = 56,533$)							
Effect	-186.62		-10.68		-118.25		-57.68
(s.e.)	(88.13)		(86.06)		(16.68)		(7.03)
%	100%		6%		63%		31%
Baseline				2,994			
% of Baseline	-6.23%		-0.36%		-3.95%		-1.93%
Panel B. Risk Premium ($n = 27,153$)							
Effect	-268.33		-116.64		-81.76		-69.93
(s.e.)	(89.45)		(87.30)		(12.18)		(5.23)
%	100%		43%		30%		26%
Baseline				2,939			
% of Baseline	-9.13%		-3.97%		-2.78%		-2.38%
Panel C. Self-Employed ($n = 39,972$)							
Effect	-457.44		-325.01		-91.35		-41.09
(s.e.)	(153.10)		(151.77)		(12.26)		(5.03)
%	100%		71%		20%		9%
Baseline				2,918			
% of Baseline	-15.67%		-11.14%		-3.13%		-1.41%
Panel D. White-Collar ($n = 39,315$)							
Effect	-161.17		10.45		-80.54		-91.08
(s.e.)	(110.21)		(105.92)		(18.13)		(8.76)
%	100%		-6%		50%		57%
Baseline				2,951			
% of Baseline	-5.46%		0.35%		-2.73%		-3.09%

NOTE.— Claims data, 2005-2008. Estimates obtained after combined approach of three difference-in-difference regressions, see text for details. Panels refer to different subgroups. Standard errors in parentheses, obtained after bootstrapping with 400 replications. *Baseline* value refers to the average claimed amount that would have occurred in the treatment group in the treatment period (2008) in the absence of treatment. Control variables in all regressions: age (linear and quadratic term), dummies for the presence of a risk premium, female gender, self-employment or white-collar workers, distinguishing plans that are only for physicians from other plans.

5.3 Health Effects

As can be seen from Table 1, two of the plans we include in our treatment group abolished the refunds scheme in 2010-2011. In this section we use these plans to study health effects from foregone utilization due to the intervention. To study whether the results may be generalized, Table 8 compares the decomposition results from Table 4 in panel A with the new decomposition results that are obtained with only two abolishing plans in panel B. For obvious reasons, the standard errors are substantially larger for the smaller sample of refund-abolishing plans but in all other aspects the two samples yield similar inference.

TABLE 8.
Comparison of main results for all and only refund abolishing insurance plans

	Before-After: Trends			**DiD**
	Common (1)	Plan (2)	Individual (3)	(4)
Panel A. All plans				
Total	-417.58 (65.14)	-417.89 (65.13)	-430.29 (70.63)	-244.10 (68.48)
Intensive Margin	-235.70 (64.67)	-235.90 (64.66)	-248.26 (70.36)	-84.78 (67.13)
Extensive Margin	-139.11 (8.35)	-139.24 (8.35)	-138.79 (8.51)	-93.40 (9.60)
Automatic	-42.78 (2.64)	-42.75 (2.65)	-43.24 (2.71)	-65.93 (3.90)
Panel B. Abolishing plans				
Total	-364.13 (175.24)	-364.07 (175.22)	-418.28 (181.05)	-241.56 (131.27)
Intensive Margin	-210.85 (164.72)	-210.80 (164.71)	-269.26 (170.11)	-129.94 (124.58)
Extensive Margin	-113.62 (35.39)	-113.61 (35.39)	-105.38 (36.32)	-65.90 (25.75)
Automatic	-39.66 (13.29)	-39.66 (13.29)	-40.77 (13.13)	-43.64 (13.59)

NOTE.— Claims data, 2005-2008. Estimates obtained after combined approach of three before-after-regressions (columns 1–3) or three difference-in-difference regressions (column 4), see text for details. Standard errors in parentheses, obtained after bootstrapping with 400 replications. Baseline value refers to the average claimed amount that would have occurred in the treatment group in the treatment period (2008) in the absence of treatment. Control variables in all regressions: age (linear and quadratic term), dummies for the presence of a risk premium, female gender, self-employment or white-collar workers, distinguishing plans that are only for physicians from other plans.

In what follows, we provide results on utilization in 2005-2007 and 2010-2011 and focus only on the two treated plans that abolish refunds in 2010-2011 and the control plans. Table 9 provides an overview on this new subsample of plans.

TABLE 9.
Health effects subsample overview

Group	Plans	Refund size				Observations
		2005-2007	2008	2009	2010-2011	
Treatment	2	0.5/1/1.5/2	3	1/3	–	1,228 (13%)
Control	6	0.5/1	0.5/1	–	–	8,302 (87%)

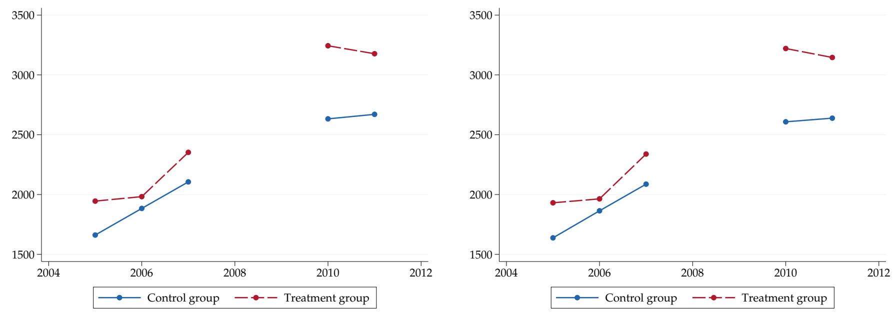
NOTE.— Refund size measured in monthly premiums if claim-free for one/two/three/four consecutive years. Listed until maximum is reached.

Figure 5 plots average claiming over time, by treatment status. The left panel (a) simply plots average individual claiming Y for both the two treated plans and the control plans. However, as D differs substantially between treatment and control group, we define Y' as a censored version of Y, according to the formula:

$$Y' = \begin{cases} 0 & \text{if } Y \leq D_0 \\ Y & \text{if } Y > D_0. \end{cases} \tag{8}$$

This censoring affects 2,753 person-year-observations in both 2010 and 2011 but does not have a strong influence on the annual average values. Accordingly, panel (b) of Figure 5 plots the evolution of Y' over time. The pattern resembles the time trends in panel (a)

FIGURE 5.
Average total claims over time by definition of Y



(a) Y

(b) Y'

NOTE.— Claims data, 2005 – 2011. (a) Y: Individual claims. (b) Y': Y censored at D_0. See text for details.

Total claims Y may be used as an outcome in a DiD-regression analogue to the baseline analysis studied so far. Table 10 reports the results from this exercise based on observations from years 2005–2007 and 2010–2011. Three different specifications are estimated, with the respective results contained in the three columns of the table. Column (1) only includes a post-treatment indicator equal to one for 2010–2011. Column (2) additionally includes yearly dummies and interactions with the treatment indicator and column (3) also controls for age (linear and quadratic term), dummies for the presence of a risk premium, being female, self-employment or white-collar workers. In correspondence with the graphical evidence the results show that claiming in the two treated plans is higher in both 2010 and 2011, hence indicating an increase in health expenditures when the refunds were abolished.[9]

TABLE 10.
Estimated effects on total claims for refund abolishing plans, 2010-2011: DiD model

	(1)	(2)	(3)
Treatment Group	209.757**	246.846	518.637***
	(102.469)	(177.440)	(174.520)
Post Treatment Period	767.867***		
	(58.163)		
Treatment × Post	349.116**		
	(162.020)		
Treatment × Year 2005		36.877	7.771
		(250.939)	(245.516)
Treatment × Year 2006		-148.144	-162.789
		(250.939)	(245.510)
Treatment × Year 2010		364.149	409.192*
		(250.941)	(245.527)
Treatment × Year 2011		259.906	319.381
		(250.941)	(245.540)
Observations	47,646	47,646	47,646

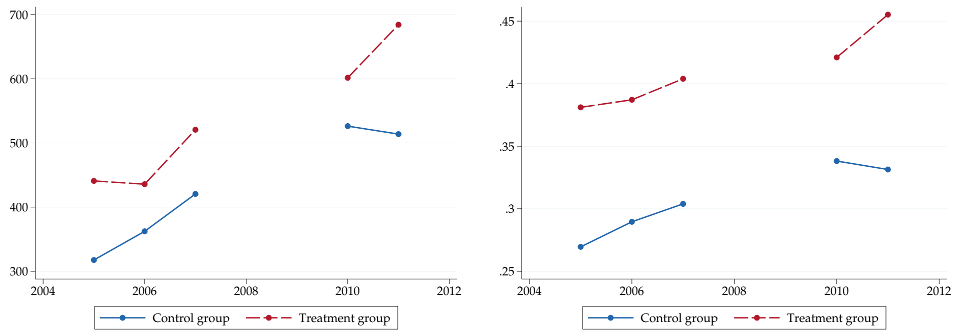
NOTE.— Claims data, 2005-2007, 2010-2011. Outcome in each DiD-regression is Y_1. Regressions differ as follows: (1) Only post-treatment indicator (equals one for 2010-2011). (2) Yearly Dummies and Interactions of those with treatment indicator. (3) As (2) but additionally controlling for age (linear and quadratic term), dummies for the presence of a risk premium, female gender, self-employment or white-collar workers. Standard errors in parentheses. * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$)

In order to see potential negative health effects we may study annual dental claims which can sometimes be seen as a non-acute investment in health that may incur significantly higher costs in the future if neglected in the present. Hence, it makes a good case for the incentives provided by the refund policy. Figure 6 plots the average dental claims and prob-

[9]The coefficients are of similar size if observations from 2010 are disregarded. Very similar results are obtained by using Y' instead of Y as the outcome variable.

ability of dental claims. The DiD results reported in Table 11 shows that there is an increase, however insignificant, in 2011 for both the amount claimed and the probability of claiming dental benefits. In sum, the results obtained point to some evidence that health care expenditures increased for individuals that were subject to the new refund policy in 2008.

FIGURE 6.
Average dental claims and probability of submitting a dental claim over time



(a) Average dental claims

(b) Probability of dental claim

NOTE.— Claims data, 2005 – 2011.

TABLE 11.
Estimated effects on dental claims for refund abolishing plans, 2010-2011: DiD model

	Average dental claims			**Probability of dental claim**		
	(1)	(2)	(3)	(4)	(5	(6)
Treatment Group	98.924***	100.071**	125.374***	0.103***	0.100***	0.117***
	(26.454)	(45.806)	(45.880)	(0.008)	(0.014)	(0.014)
Post Treatment Period	153.171***			0.047***		
	(15.014)			(0.005)		
Treatment × Post	23.926			0.000		
	(41.825)			(0.013)		
Treatment × Year 2005		23.176	24.989		0.012	0.011
		(64.779)	(64.544)		(0.020)	(0.020)
Treatment × Year 2006		-26.648	-25.776		-0.002	-0.003
		(64.791)	(64.554)		(0.020)	(0.020)
Treatment × Year 2010		-24.787	-27.224		-0.017	-0.016
		(64.779)	(64.546)		(0.020)	(0.020)
Treatment × Year 2011		70.348	67.051		0.024	0.025
		(64.780)	(64.550)		(0.020)	(0.020)
Observations	47,642	47,642	47,642	47,642	47,642	47,642

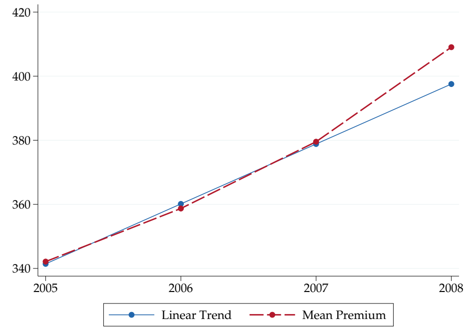
NOTE.— Claims data, 2005-2007, 2010-2011. Outcomes: Annual sum of dental claims per client (columns 1–3), indicator if client submitted dental claims in this year or not (columns 4–6). Regressions differ as follows: (1) Only post-treatment indicator (equals one for 2010-2011). (2) Yearly Dummies and Interactions of those with treatment indicator. (3) As (2) but additionally controlling for age (linear and quadratic term), dummies for the presence of a risk premium, female gender, self-employment or white-collar workers. Standard errors in parentheses. * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$)

5.4 Insurer Profits

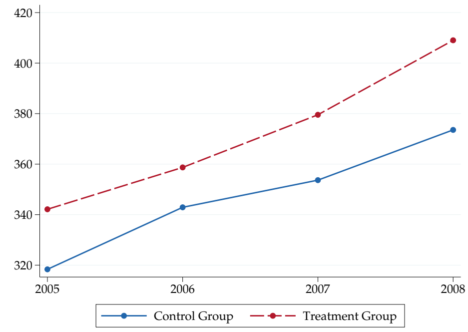
Replacing the client outcomes by the insurer's revenues and costs in our analysis may tell us something about the effects on the profits of the insurer from implementing the new refund policy. Specifically, Figure 7 illustrates how using premiums, actual reimbursement and paid out refunds evolve over time. Panel (a) depicts how the average premium evolved over time in the treatment group in comparison to a linear trend, while panel (b) shows the same evolution for the DiD analysis. Analogously, panel (c) and (d) and (e) and (f) depicts the same for the average reimbursed amount per client and paid out refunds, respectively. All three outcomes seem to have changed at the time of the implementation of the refund policy.

FIGURE 7.
Changes in insurer's revenues over time

Average insurance premium

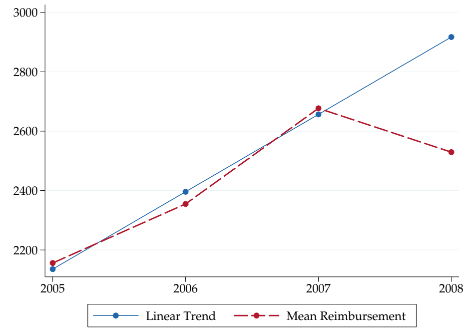


(a)
Before-After

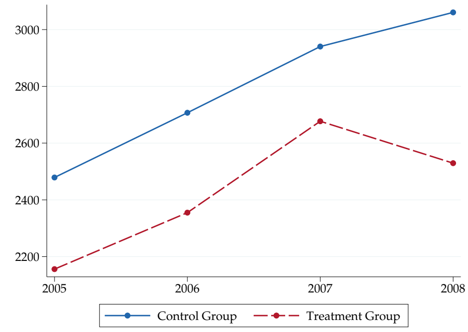


(b)
DiD

Average claim reimbursement

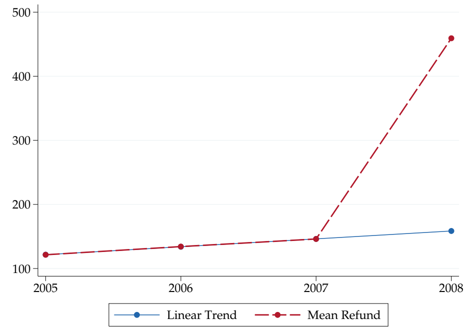


(c)
Before-After

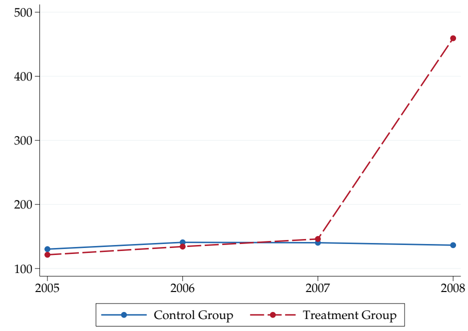


(d)
DiD

Average refund



(e)
Before-After



(f)
DiD

NOTE.— Claims data, 2005 – 2011.

Table 12 provides the corresponding regression estimates for the changes in the three revenue components for each empirical specification. The main difference between the two models is due to the higher change in reimbursement from the before-after model which is almost doubled compare to the DiD model. According to the DiD-results reimbursement only decreased by about €236, while the estimated effects for premiums and refunds are of similar absolute size as according to before-after approach. Favoring the results from the DiD model, the interventions seems not to have been profitable for the insurer as they lost $9 + 236 - 316 = -$ €71 on average per client.

TABLE 12.
Estimated changes in insurer's revenues, 2010-2011

	Premium (1)	Reimbursement (2)	Refund (3)
Panel A. Before-after model $(n = 48,114)$			
Effect	11.69***	-398.72***	301.43***
(s.e.)	(0.90)	(56.64)	(3.15)
Observations	192,456	192,456	192,456
Panel B. DiD model $(n = 83,686)$			
Effect	9.18***	-235.50***	315.53***
(s.e.)	(1.23)	(68.09)	(3.04)
	334,744	334,744	334,744

NOTE.—Claims data, 2005 – 2011. All models controls for yearly dummies and interactions of those with treatment indicator, age (linear and quadratic term), dummies for the presence of a risk premium, female gender, self-employment or white-collar workers. Standard errors in parentheses. * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$)

6 Conclusion

Policy makers and academics are concerned with rising costs in the health care sector. Health insurance markets are particularly important given the problematic consequences of moral hazard and adverse selection. A broad body of literature aims to understand how different characteristics of a health insurance plan affect individual behavior, which is key for the optimal design of such plans that incorporates cost-containment appropriately. This paper adds to this literature and provides evidence that individuals insured in the German private health insurance market indeed seem to react upon incentives to reduce health care with respect to premium refunds.

Premium refunds are monetary incentives for the insured to reduce unnecessary health care spending. Contingent on being claim-free for a whole calendar year a client receives a multiple of monthly premiums back as a refund in the following year. We use rich data from a large German private health insurer and exploit a policy shift that increased the refund size for some insurance plans while leaving other plans unchanged. Applying a difference-in-differences design we find that clients in plans that were subject to a increase in the refund significantly reduced their claims by on average 8% (€ 240) compared to other clients. Decomposing the effect into an automatic, an extensive and an intensive margin effect, we find that the policy reduced claiming for all three components but in particular the propensity to claim on the extensive margin. Studying effect heterogeneity we find that females, older and risk-rated clients reacted more to the new refund policy. Furthermore, the intervention seems not have been profitable for the insurer as our estimates suggests that increased revenues from lower reimbursement claims were lower than increased costs from payout of refunds. Finally, we find some evidence of health effects as utilization increased after the refunds scheme was abolished.

In summary, individuals seem to responding to changes of the design of health insurance plans with respect to premium refunds. In line with the previous literature we confirm that individuals seem to understand the dynamic and non-linear design of insurance plans and react in line with general intuition and economic theory. This is indicating that unnecessary individual health care costs might indeed be reduced by such demand side incentives. These insights might deliver useful policy implications for the design of health insurance contracts and respective approaches to moral hazard in such settings.

References

ARON-DINE, A., EINAV, L. and FINKELSTEIN, A. (2013). The RAND Health Insurance Experiment, Three Decades Later. *Journal of Economic Perspectives*, **27** (1), 197–222.

—, —, — and CULLEN, M. (2015). Moral Hazard in Health Insurance: Do Dynamic Incentives Matter? *Review of Economics and Statistics*, **97** (4), 725–741.

BROT-GOLDBERG, Z. C., CHANDRA, A., HANDEL, B. R. and KOLSTAD, J. T. (2015). *What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics*. Working Paper 21632, National Bureau of Economic Research.

BUNDORF, M. K., LEVIN, J. and MAHONEY, N. (2012). Pricing and welfare in health plan choice. *The American Economic Review*, **102** (7), 3214–3248.

CABRAL, M. (2016). Claim timing and ex post adverse selection. *The Review of Economic Studies*, p. rdw022.

CARLIN, C. and TOWN, R. (2010). *Adverse selection: The dog that didnt bite*. Working Paper.

DALTON, C. M. (2014). Estimating demand elasticities using nonlinear pricing. *International Journal of Industrial Organization*, **37**, 178–191.

—, GOWRISANKARAN, G. and TOWN, R. (2015). *Myopia and Complex Dynamic Incentives: Evidence from Medicare Part D*. Working Paper 21104, National Bureau of Economic Research.

EINAV, L. and FINKELSTEIN, A. (2011). Selection in insurance markets: Theory and empirics in pictures. *The Journal of Economic Perspectives*, **25** (1), 115–138.

—, —, PASCU, I. and CULLEN, M. R. (2012). How general are risk preferences? choices under uncertainty in different domains. *The American Economic Review*, **102** (6), 2606–2638.

—, — and SCHRIMPF, P. (2015). The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D. *The Quarterly Journal of Economics*, **130** (2), 841–899.

ELLIS, R. P. and ZHU, W. (2016). Health plan type variations in spells of health-care treatment. *American Journal of Health Economics*.

KARLSSON, M., KLEIN, T. and ZIEBARTH, N. R. (2016). Skewed, Persistent and High before Death: Medical Spending in Germany. *Fiscal Studies*, **forthcoming**.

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

MANNING, W. G., NEWHOUSE, J. P., DUAN, N., KEELER, E. B. and LEIBOWITZ, A. (1987). Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. *The American Economic Review*, **77** (3), 251–277.

PKV (2015). *Zahlenbericht der Privaten Krankenversicherung 2014*. Report, Cologne, Germany.

ZWEIFEL, P. (1987). Bonus systems in health insurance: a microeconomic analysis. *Health Policy*, **7** (2), 273–288.

— (1992). *Bonus Options in Health Insurance*. Dordrecht, The Netherlands: Springer Science & Business Media.

Appendix: Additional tables and figures

TABLE A.1.
Descriptive statistics, 2005

	Treated		Controls	
	mean	sd	mean	sd
Female	0.11	0.31	0.30	0.46
Age	41.08	7.68	45.90	10.14
Premium	342.13	123.41	318.36	152.11
Deductible	185.50	150.87	1070.47	131.12
Risk premium	0.32	0.47	0.33	0.47
Self employed	0.32	0.47	0.69	0.46
White-collar	0.65	0.48	0.23	0.42
Other occupation	0.03	0.18	0.08	0.27
Observations	48,114		35,572	

NOTE.— Claims data, 2005 – 2008. All monetary variables measured in 2011 euros.

TABLE A.2.
Descriptive statistics, 2006

	Treated		Controls	
	mean	sd	mean	sd
Female	0.11	0.31	0.30	0.46
Age	42.08	7.68	46.90	10.14
Premium	358.70	134.84	342.91	167.06
Deductible	182.73	148.62	1054.51	129.16
Risk premium	0.32	0.47	0.33	0.47
Self employed	0.32	0.47	0.69	0.46
White-collar	0.65	0.48	0.23	0.42
Other occupation	0.03	0.18	0.08	0.27
Observations	48,114		35,572	

NOTE.— Claims data, 2005 – 2008. All monetary variables measured in 2011 euros.

TABLE A.3.
Descriptive statistics, 2007

	Treated		Controls	
	mean	sd	mean	sd
Female	0.11	0.31	0.30	0.46
Age	43.08	7.68	47.90	10.14
Premium	379.55	143.63	353.67	176.82
Deductible	178.36	145.10	1030.37	126.20
Risk premium	0.32	0.47	0.32	0.47
Self employed	0.32	0.47	0.69	0.46
White-collar	0.65	0.48	0.23	0.42
Other occupation	0.03	0.18	0.08	0.27
Observations	48,114		35,572	

NOTE.— Claims data, 2005 – 2008. All monetary variables measured in 2011 euros.