Journal of Public Economics 230 (2024) 105061

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

No-claim refunds and healthcare use[☆]

Daniel Avdic [a,*], Simon Decker [b], Martin Karlsson [c], Martin Salm [d]

[a] Department of Economics, Deakin University, 70 Elgar Road, Burwood VIC 3125, Australia
[b] RWI Essen, Germany
[c] CINCH, University of Duisburg–Essen, Germany
[d] Department of Econometrics and Operations Research, Tilburg University, Netherlands

ARTICLE INFO

JEL classification:
D82
G22
I13

Keywords:
No-claim refund
Moral hazard
Health insurance

ABSTRACT

No-claim refunds are cost-control instruments which stipulate a payback agreement contingent on one or more claim-free years. We study how such no-claim refunds affect claiming behavior using claims data from a large German health insurer and a policy that increased the refund size for certain plans. We propose a method to decompose the effect on claims into behavioral and non-behavioral components, and show that individuals responded to the refund policy by reducing claims by eight percent on average. The effect persisted for several years; behavioral responses were stronger for clients with more to gain from the policy; and reductions in claims were not restricted to treatments of questionable medical value.

1. Introduction

Patient cost-sharing is an important tool to contain high and increasing healthcare expenditures in many countries. On average across OECD countries, patients pay around 20 percent of total healthcare expenditures out-of-pocket (OECD, 2015). An economic justification for patient cost-sharing is that it can reduce moral hazard and inefficient outcomes (Pauly, 1968). However, this justification is critically dependent on the assumption that clients understand possibly quite complex cost-sharing schemes in insurance contracts.

Patient cost-sharing in health insurance contracts can take many different forms. The most common form of cost-sharing are annual deductibles in which clients pay out-of-pocket for their healthcare use up to the deductible limit, and the insurance provider pays for any costs above the limit. Other forms of cost-sharing include no-claim refunds, co-payments as a constant percentage of costs, or more complex nonlinear price schedules such as the donut hole for Medicare Part D in the United States. Previous studies have found that the response to different forms of cost-sharing incentives can vary (see, e.g., Remmerswaal et al., 2019; Hayen et al., 2021), and clients might not fully take into account the dynamic incentives inherent in nonlinear price schedules (see, e.g., Aron-Dine et al., 2015; Dalton et al., 2019; Einav et al., 2015; Brot-Goldberg et al., 2017; Abaluck et al., 2018). It is therefore important to examine in more detail how clients respond to alternative forms of cost-sharing that have received less attention in the literature.

In this paper we contribute to the literature on consumer behavior in health insurance markets by studying no-claim refunds, a largely unexplored cost-control instrument widely used by health insurance providers in the German private health insurance market.[1] No-claim refunds are used as a monetary incentive for clients to withhold insurance claims. Specifically, contingent on being claim-free for a full calendar year, a client with a refund option stipulated in their insurance contract is repaid a proportion of last year's annual paid-in insurance premium.

We use unique claims data from a large German private health insurer to empirically study how the option of receiving a no-claim refund affects clients' claiming behavior. To alleviate empirical concerns from clients sorting into insurance contracts based on underlying preferences and needs for healthcare consumption, we exploit an insurer policy that unexpectedly increased the refund size for some insurance plans,

☆ We thank the editor Sarah Miller, three anonymous referees, John Buckell, Burkhard Hehenkamp, Maryna Ivets, 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 Bergen, Essen Health Conference 2017, EWEHE-Prague, Heidelberg, Hohenheim, IAAE-Sapproro, Lund University, NBER Summer Institute 2017, RWI-Essen and University of Gothenburg. We gratefully acknowledge institutional funding for CINCH from the Federal Ministry of Education and Research, grant number 01EH1102A.

* Corresponding author.
E-mail address: d.avdic@deakin.edu.au (D. Avdic).

1 Zweifel (1987) and Zweifel (1992) has previously studied the impact of no-claim refunds in the German private health insurance system, finding that such incentives are associated with reduced healthcare costs. However, these findings may to some extent be driven by selection of healthier individuals into contracts with refund options which cannot be ruled out in the author's data.

https://doi.org/10.1016/j.jpubeco.2023.105061
Received 19 August 2022; Received in revised form 15 December 2023; Accepted 20 December 2023
Available online 10 January 2024

while refunds in other contracts remained unchanged, in a generalized difference-in-differences framework. To distinguish between behavioral and non-behavioral responses to the policy, we also propose a novel method to decompose the overall impact of the insurance policy into an intensive margin effect (total amount claimed), an extensive margin effect (propensity of claiming) and an ''automatic'' effect (mechanical impact of raising the co-payment threshold). The distinction between the two former and the latter factors is important as we are mainly interested in clients' behavioral responses to the refund policy.

We find that the refund policy reduced the average annual amount claimed by clients by eight percent (€200) in the year the policy was enacted. Our decomposition approach further reveals that this effect is mainly driven by the extensive and automatic effects, with the former constituting approximately two thirds of the total impact on claims. As expected from the policy design, we find no effect on the intensive margin (i.e., the average size of positive claims). Using an event study model to study effect dynamics in subsequent years, we find that the initial impact persisted for several years after the policy was introduced, even as the refund scheme became less generous after the first year and economic incentives to claim later refunds were lower.

To understand the mechanisms behind clients' responses to the refund policy, we extend our analysis by studying subgroups of clients based on their expected healthcare utilization using Adjusted Clinical Groups (ACG) scores.[2] ACG scores can be used to approximate the heterogeneous incentives that the refund policy provided to clients by quantifying their expected healthcare expenditures. We find that clients with ACG scores just below the mean responded relatively more to the policy compared to individuals with very high or very low scores. This result is in line with our predictions, as it is unlikely that, for example, chronically ill clients would be able to benefit from the policy even if they wanted to. Likewise, clients with very low ACG scores are unlikely to submit any claims regardless of the refund policy.

We further conduct an analysis of heterogeneous effects based on claim types, utilizing diagnosis- and category-specific information to investigate whether clients systematically favored specific treatment types to qualify for the no-claim refund. Our empirical findings are mixed in that the estimates provide no consistent pattern with respect to typical patient-initiated discretionary treatments, including dental care, visual aids, and alternative medicine. However, our results do suggest a significant reduction in care for non-specific diagnoses, which serve as another indicator of discretionary care in our context. Taken together, these results suggest that clients curtailed a broad spectrum of services in response to the economic incentives stemming from the no-claims refund policy.

Previous studies have pointed out several dimensions where the incentives generated by no-claim refunds differ from incentives in other forms of patient cost-sharing. First, responses to no-claim refunds and deductibles can differ if clients are loss averse (Kahneman and Tversky, 1979). Clients might view the failure to secure a no-claim refund as a forfeited gain, while they see a deductible payment as a direct loss.[3] In this case, loss-averse clients would prefer a no-claim refund over a deductible and react less to a no-claim refund than to a deductible payment of the same nominal size (Johnson et al., 1993). These predictions are confirmed by two recent empirical studies examining premium refunds in the Netherlands (Remmerswaal et al., 2019; Hayen et al., 2021). In contrast to these studies, we perform an in-depth examination of no-claim refunds rather than comparing clients' responses to no-claim refunds and deductible payments.[4]

Second, clients might reduce their use of different types of care in response to deductibles and refunds. Kairies-Schwarz et al. (2023) find such different responses in a discrete choice experiment. This motivates examining patients' responses to no-claim refunds by care type in a field setting. Our results confirm findings for deductibles suggesting that patients cut down on a broad spectrum of services not confined to treatments of questionable medical value (Manning et al., 1987; Brot-Goldberg et al., 2017).

Third, no-claim refunds differ from other forms of patient cost-sharing since they provide solely dynamic incentives rather than a combination of both dynamic and current incentives.[5] Our finding that individuals respond to incentives when they are framed as a deferred reward therefore contributes to the literature on how individuals respond to dynamic incentives inherent in many cost-sharing schemes (see, e.g., Aron-Dine et al., 2015; Dalton et al., 2019; Einav et al., 2015; Brot-Goldberg et al., 2017; Abaluck et al., 2018). The result that more incentivized individuals responded more strongly to the policy reinforces the interpretation that individuals do take future rewards into account when making decisions about current healthcare use and contrasts findings by Brot-Goldberg et al. (2017) who show that clients respond to an increase in a deductible irrespective of how exposed they are to the reform. An explanation for the mixed findings could be that individuals respond differently to dynamic incentives imposed by no-claim refunds and deductibles.

Finally, with respect to healthcare policy our findings suggest that no-claim refunds constitute a viable alternative to deductibles for insurers. No-claim refunds achieve cost-sharing goals, but without some of the well-known disadvantages of deductibles. Specifically, they provide less challenges for liquidity-constrained clients, and they might face less resistance from clients who dislike deductibles (Johnson et al., 1993; Bhargava et al., 2017). However, one important disadvantage of deductibles also applies to no-claim refunds: reductions of claims include treatments that are likely of high medical value. Although our results do not indicate any immediate health effects from neglecting essential care in the years following the policy change, the finding that some common elective healthcare services were impacted by the policy suggests that clients may face increased healthcare expenses later in life.

2. Institutional setting and context

2.1. *The German healthcare system*[6]

In Germany, health insurance can be obtained either through statutory health insurance (SHI) or through private health insurance (PHI). While enrolling in SHI is compulsory for the majority of the German population, certain groups, including civil servants, the self-employed, freelancers, and high-income earners, may opt out of SHI to join a PHI provider. In particular, individuals with taxable earnings above the annual income threshold *(Versicherungspflichtgrenze)* applying, which amounts to €64,350 in 2021, are allowed to opt out of SHI. In 2019, roughly 11 percent of the German population was insured through PHI as their primary source of insurance coverage (AGPHI, 2020).[7]

---

[2] See, e.g., Brot-Goldberg et al. (2017) for a recent application of ACG scores in economics and https://www.hopkinsacg.org/ for further details on the ACG system. Our analyses are based on The Johns Hopkins ACG® System Version 11.1.

[3] This reasoning applies if clients consider their wealth after paying insurance premiums as reference point. Van Winssen et al. (2016) discuss clients' responses to no-claim refunds for alternative reference points.

[4] The no-claim refunds examined in our study also differs from the refund used in the Netherlands. In the Netherlands, individuals could still receive a partial refund if they have positive but low healthcare expenditures, while the refund considered in our study is a true no-claim refund.

[5] For a definition of both dynamic and current incentives see e.g. Klein et al. (2022).

[6] We refer to Simon (2017) and Blümel et al. (2020) for a more detailed description of the German healthcare system.

[7] In addition, some individuals in the SHI system purchase supplemental PHI coverage.

Insurance premiums vary between the two systems. In SHI, premiums are set by the Federal Ministry of Health. In 2021, premiums are 14.6 percent of earnings, which are paid in equal shares by employers and employees. Premiums cover a fixed set of services described in the German Social Law. In contrast, PHI premiums are based on individual contractual agreements between the insurance company and the client which outline the set of covered services and the percentage of coverage, and they are adjusted for the person's health risk and age of entry into the private system. PHI contracts determine a bundle of covered services and reimbursement rates from which the individual can flexibly choose to include or exclude elements, and premiums are adjusted according to the chosen benefit package. The PHI system is regulated by the Federal Ministry of Health to ensure that the insured do not face large premium increases as they age and are not overburdened if their income decreases (cf., Atal et al., 2020).

Individuals in both systems have free choice among GPs and specialists. Registration with a family physician is not required, and GPs have no formal gate-keeping function. Under SHI, GPs and specialists are generally reimbursed on a fee-for-service (FFS) basis according to a uniform fee schedule that is negotiated between sickness funds and regional associations of physicians. For private patients, GPs and specialists are also paid on a FFS basis, but private tariffs are usually higher than the tariffs in the SHI fee schedule (Jürges, 2009). Inpatient care is paid per admission through a system of DRGs, which are revised annually. DRGs cover all services and all physician costs. All drugs, both patented and generic, are placed into groups with a reference price serving as a maximum level for reimbursement, unless an added medical benefit can be demonstrated. For new drugs with added benefit, the Federal Association of Sickness Funds negotiates a reimbursement price, based on the manufacturer's price, that is applied to all patients (i.e., both privately and publicly insured). For PHI, providers send medical bills directly to patients. Patients pay upfront and submit claims to the insurance company for reimbursement subject to the specific cost-sharing arrangement (see, e.g., RKI, 2015).[8]

2.2. No-claim refunds

An important aspect of the German PHI system that distinguishes it from the SHI system is its cost-sharing arrangements, mainly consisting of deductibles and no-claim refunds. Deductibles are standard in the PHI system implying that clients pay a certain amount of healthcare costs out-of-pocket before insurance coverage is activated. In practice, since PHI clients pay their medical costs upfront and then seek reimbursement from their insurer, any bills submitted to the insurer with monetary claims that amount to a sum below the deductible will not be reimbursed.

Furthermore, many insurers offer no-claim refunds to their clients as financial incentives to limit the use of non-essential healthcare. A no-claim refund option entitles the client to receive a share of the previous year's total paid-in insurance premium refunded by the insurer, contingent on being claim-free for the entire calendar year. The client is instantly disqualified for a refund once a claim for a particular calendar year has been submitted to the insurer, even if the claim is not reimbursed. If the client did not submit any claims, the insurer will automatically transfer the contracted amount of the refund to the client at a specified date in the following year.[9] Importantly, the combination of a deductible and a no-claim refund option suggests that it is rational for clients to submit their claims to the insurer only when the sum of their medical bills in one calendar year exceeds the sum of the deductible and the refund combined. Clients make payments directly to healthcare providers and then choose whether to keep the bills or submit them to the health insurer for reimbursement.

The insurance policy we analyze in this paper is a sudden and unexpected increase in no-claim refunds for selected insurance plans from a large German PHI provider. The change was announced in February 2008 and applied to claims made during the calendar year of 2008. Before the refund policy changed in 2008, an experience-rated refund was offered to all plans. Each additional year that a client withheld from submitting any medical claims to the insurer was rewarded with a refund amounting to 0.5 monthly insurance premiums, up to a maximum of two monthly premiums. In 2008, the refund structure changed and every client instead received a refund of three monthly premiums if no claims were filed, regardless of their previous claiming history. The stated reason for this change was that the insurance provider acted in response to a bill by the German legislature that enabled the portability of old-age provisions to other insurers in the German PHI sector. The bill meant that clients no longer had to risk losing their accumulated reserve provisions when switching insurance providers. In response, the insurer increased the generosity of the no-claim refunds for selected plans which were most affected by the reform to retain its clients.[10],[11]

3. Conceptual framework

Fig. 1 presents a stylized illustration of how a change in the no-claim refund scheme affects the budget set available to clients. The figure shows possible combinations of consumption of healthcare (M) and other commodities (C), before and after the change in policy. For simplicity, we assume that there was no refund scheme prior to the intervention, and we normalize all prices to one.

Prior to the intervention the client pays the full cost of care up to the level of the deductible D_0, after which healthcare costs are fully reimbursed. For a client with zero healthcare costs, the intervention is equivalent to an increase in the client's income, I, by δR, where R is the refunded monetary amount they can expect after a claim-free year, and $\delta \in [0,1]$ is the individual's time discount factor, reflecting the fact that the no-claim refund is paid out several months after the end of the calendar year. Clearly, $\delta R = D_1 - D_0$. Moreover, the shape of the budget constraint imposed by a no-claim refund in Fig. 1 is very similar to the shape of the budget constraint imposed by the deductible D_0.

The rationality of consumers in the presence of dynamic incentives is a hotly debated topic in the literature examining patients' responses to cost-sharing incentives. 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 start by introducing 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 health status H. We assume that the function is concave whenever H is below perfect health. When $H = H^{max}$ the marginal utility with respect to healthcare consumption is zero. We also assume that whenever the individual is not in perfect health ($H < H^{max}$), their marginal rate of substitution is greater than the price of healthcare services when the amount consumed approaches zero:

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

[8] Inpatient services are typically reimbursed directly between hospitals and the insurer due to the high costs encountered here.

[9] Some services, such as certain types of inpatient or preventive care, may be exempt from the refund policy and clients may claim reimbursement for such services without forgoing the refund option. However, such exemptions only account for a very small share of claims (1.5 percent) in our analysis sample.

[10] Atal et al. (2017) document that, in contrast to the insurer's concerns, external exits did not increase in response to the portability-reform.

[11] We provide further details about the reform and the motivation of the insurer to change no-claim refunds in online Appendix A.

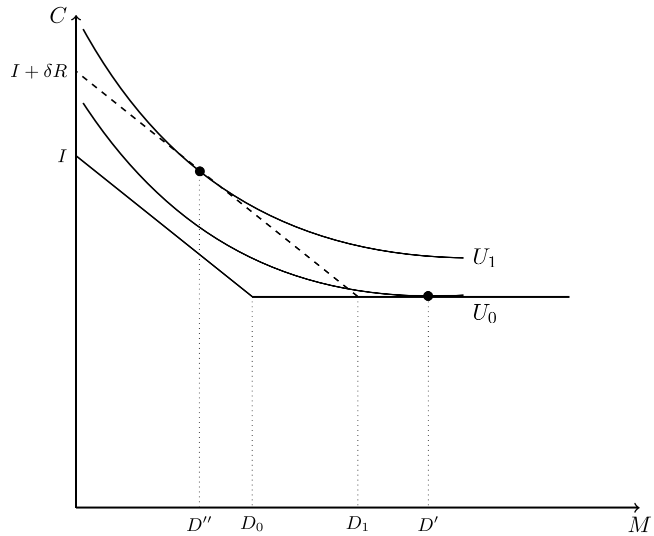


Fig. 1. Impact of a hypothetical no-claim refund increase. Note.— Expected effects of an increase in no-claim refunds R on consumption C as a function of medical expenditures M for a consumer with utility function $U(M, C \mid H)$ and two hypothetical indifference curves U_0 and U_1. I is disposable income and δ is the time discount factor. D_0 and D_1 are pre- and post-policy deductible and potential refund levels and D' and D'' are optimal medical expenditures before and after the policy was introduced, respectively.

Based on these elementary assumptions, it is possible to produce some useful comparative statics regarding how rational consumers can be expected to respond to a change in no-claim refunds. One prediction follows immediately from an inspection of the budget constraint in Fig. 1: 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.

In addition, our assumption in Eq. (1) above regarding the MRS as M approaches zero allows us to rule out another possibility: no consumer with healthcare utilization $M \in (0, \infty)$ will find it optimal to switch to $M = 0$ in the presence of no-claim refunds. Conversely, no consumer with $M = 0$ to begin with will find it optimal to change their healthcare utilization either, since they are assumed to be in perfect health, and thus they do not derive utility from healthcare consumption. Consumers with utilization $M \in (0, D_0)$ under the old regime will *increase* their healthcare consumption, which will be within the segment $(0, D_1)$: this reflects a pure income effect from the increase in no-claim refunds. Consumers with utilization in the region $M \in (D_0, D_1)$ may move to any point in the segment $(0, D_1)$: they experience both an increase in their incomes and an increase in the relative price of healthcare. Finally, consumers who are initially in the segment $M > D_1$ will either stay were they are in case they are in relatively poor health so that the utility loss of decreasing their healthcare utilization to below D_1 is large, or they might shift consumption to the segment $(0, D_1)$ if they are in relatively good health with initial consumption not far above D_1.[12]

Thus, based on some basic assumptions with respect to the preferences of consumers, we may formulate a set of hypotheses regarding the reactions to an increase in no-claim refunds:

1. The probability of observing expenditure $M > D_1$ should decrease.
2. Average expenditures above the new threshold D_1 should stay constant or increase.
3. The proportion of positive expenditure (as opposed to zero expenditure) below the new threshold will increase.

As will be outlined below, hypotheses (1) and (2) relate to directly observable facts. Hypothesis (3) relates to quantities that cannot be directly observed in the data. Nevertheless, hypothesis (3) is useful in order to make predictions about how the entire distribution of healthcare utilization will adjust. In particular if we find empirical evidence consistent with hypotheses (1) and (2), we can use hypothesis (3) to draw preliminary conclusions regarding how unobserved utilization below D_1 changes.

These predictions are derived from a static framework with rational consumers and well-behaved preferences. In this framework, incentives for no-claim refunds and deductibles are very similar, and our predictions hold both for responses to no-claim refunds and to deductibles. In the following, we discuss how the predictions of our model depend on underlying assumptions and how a violation of these assumptions can lead to different predictions for no-claim refunds and deductibles.

First, we assume that individuals take future refunds into account in their healthcare choices, e.g. $\delta > 0$. Completely myopic patients would not respond to no-claim refunds, but they would still respond to deductible payments, since no-claim refunds are reimbursed only in the future while deductible payments are due instantaneously.

Second, we abstract from liquidity concerns. Zweifel (1992) points out that a refund scheme offers the client some degree of consumption smoothing, or rather utility smoothing, as they can temporally dissociate the financial risk from the health risk. We expect that liquidity constraints can lead to less healthcare use under deductibles, but less so under no-claim refunds.

Third, we also abstract from loss-averse preferences. In our model, we make minimal assumptions on preferences. These assumptions are not necessarily violated in the presence of loss aversion. However, the marginal utility of consumption for loss-averse patients can be higher under a deductible than under a no-claim refund which implies that patients reduce healthcare use more in response to deductibles than in response to no-claim refunds.[13]

Fourth, sophisticated insurance clients could optimize the *timing* of their healthcare utilization. For example, having utilization $M' > 2D_1$ in one year and zero in the next year, is clearly preferable over having utilization equal to $M'/2$ in both years, as long as the bunching of utilization in one year does not have too severe consequences for the individual's health (Cabral, 2017). Such incentives exist for both no-claim refunds and for deductibles.

Finally, our analysis data consists only of medical expenditures the clients' have claimed reimbursement for and not medical expenditures generally. This gives rise to a number of empirical challenges that we discuss in the next section.

4. Empirical framework

4.1. Definitions

The primary variable of interest in our analysis is total annual healthcare expenditure. Denote by M_{it} the annual expenditure of 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 cumulative distribution function of such a distribution by $F(M)$. The intervention we are considering is an increase in the financial incentive to submit zero medical claims in a calendar year. For an evaluation of the intervention, it would be of interest to compare how the distribution $F(M)$, or any statistic based

[12] Table B.1 in the online Appendix provides an overview of all the changes in consumption that rational consumers would make in response to the policy.

[13] This prediction holds if the reference point is equal to the wealth after paying insurance premiums as pointed out by Johnson et al. (1993).

on it, 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 data we use consists of insurance claims, and therefore M_{it} is only observed when medical expenditures are claimed from the insurer. The observed annual claims Y are a function of actual utilization: $Y(M)$. Whenever the contract entails a deductible and/or a no-claim refund, claims tend to be censored. This censoring has a number of consequences. First, the averages forming the components in Eq. (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 refund schemes. This represents a problem since the intervention we consider is a change in no-claim refund generosity. We show below that it is nevertheless possible to draw meaningful inference based on parts of the distribution of M that are unaffected by the censoring.

We denote by D_{it} the sum of the deductible (that applies in a certain contract) and the no-claim refund (that applies after a claim-free year).[14] D_{it} determines whether it is rational to submit claims for healthcare expenditure at the end of the year. Consider for example an individual who has a deductible of €500 and a potential no-claim refund of €300. If annual healthcare expenditure amounts to €700, they will get €200 reimbursed if they submit their bills, or, alternatively, a no-claim refund of €300 if they do not. For total annual healthcare expenditures of €900 it would, on the other hand, be rational to submit the bills, at least in the absence of dynamic incentives. Thus, we write the claiming function as:

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

This claiming function implies strong assumptions. First, we assume that consumers make the decision to submit claims rationally. Second, we assume that the decision to submit claims is not affected by hassle costs, which would tend to induce insurance holders to rationally avoid submitting claims even when they would financially benefit from it. Third, we assume that there is no uncertainty in the enrollee's claim submission problem, and enrollees know that their claims will be reimbursed. Fourth, we assume that clients are not liquidity constrained. Nevertheless, we believe that this claiming function is reasonable in our context. We discuss possible violations of these assumptions and their consequences for the interpretation of our results in Section 6.4.

The intervention we consider is an unexpected increase in D_{it} that occurred in 2008: the deductible part of D_{it} remained constant whereas the no-claim refund increased by one or more monthly premiums. We can define the causal effect of this intervention on annual claims 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 the superscript in ATT^y highlights that it applies to claims only, and the superscript in $Y(M)$ indicates treatment status. Even though the effect presented in Eq. (4) undoubtedly represents a causal effect, namely by how much the intervention would change the total claims made by clients in a given contract, it is of limited interest from an economic point of view. Specifically, ATT^y combines two different responses to the intervention: an *automatic* effect arising because clients now claim according to function $Y^1(m)$ due to the higher no-claim refund applying, and a *behavioral* (or moral hazard) effect arising because the new incentives have moved the distribution of utilization from $F^0(m)$ to $F(m)$.

Rearrangement of Eq. (4) allows us to disentangle the two effects:

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

where $\mathbb{E}_1$ ($\mathbb{E}_0$) represents expectation taken in the presence (absence) of treatment. The overall effect on claims, ATT^y, now consists of three parts: an *intensive* margin (INT), capturing by how much claims above the new no-claim refund and deductible level change on average; an *extensive* margin (EXT), representing the probability of reaching the new level D_1; and an *automatic* component (AUT), reflecting expenditures between D_0 and D_1 that will not be claimed anymore.[15] The distinction between automatic and utilization effects is necessary in order to make statements about moral hazard. The distinction between the intensive and the extensive margin is also useful since it directly corresponds to hypotheses (1) and (2) presented in Section 3.

4.2. Identification

Our proposed method to separate the total effect on claims into automatic and behavioral changes solves some of the problems related to the censoring in claims. However, we also need to deal with the issue 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. This counterfactual refers to the claims that would have been submitted by treated clients if the no-claim policy had not been implemented. In order to solve this identification problem, we employ a difference-in-differences (DID) design where we define and use a control group to impute the missing counterfactual. Specifically, our control group consists of clients enrolled in a set of comparable insurance plans that were not subject to an change in no-claim refunds in 2008.

Since our aim is 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:

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

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 $M > D_0$.

Identification of the effects is contingent on three component-specific common time trend assumptions, corresponding to the INT, EXT, and AUT effects defined in the previous subsection, and one overall common time trend assumption. The latter may be defined as:

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

where D_c denotes the sum of deductible and no-claim refund that applies in the control group (and which does not change over time).

[14] 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.

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

It should be obvious from Eq. (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:

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

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

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

for the extensive margin effect (EXT) and as

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

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.

Hence, the common time trend assumptions needed for identification are somewhat stronger than those required in standard DID designs, but substantially weaker than the assumptions required for invariance to functional form as shown by Roth and Sant'Anna (2023). In the latter case, the invariance property requires that the cdf of the untreated outcome exhibits parallel trends at any point in the support.

4.3. Estimation

The overall effect of the intervention on total annual claims – ATT^y defined in Eq. (4) – may be estimated using standard DID methodology. 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 a number of plan and personal characteristics, such as plan type, age, and gender. Under the assumptions stated above, this regression would identify ATT^y or the total effect of the policy on annual claims. Formally,

$$Y_{it}^{q} = \alpha_{c_i} + \mu_t + \beta P_i + \gamma^q \left\{ \mathbb{1}(year_t > 2007) \times P_i \right\} + X_{it}'\beta_X + \varepsilon_{it} \tag{DD}$$

where q is the specific outcome margin of interest (see below), α_{c_i} and μ_t are plan and year fixed effects, and $P \in \{0,1\}$ is equal to one if client i was insured with a plan that was subject to the no-claim refund policy in 2008. Furthermore, X_{it} is the vector of control variables, and ε_{it} an error term which has mean zero if the identifying assumptions are satisfied.[16] The parameter of interest is γ^q, which represents the average impact of the no-claim refund policy on the annual amount claimed for outcome margin q.

In order to identify *behavioral* changes, we generate the new variable Y_{it}^{beh}, which is censored at the new refund and deductible level D_1 according to the following formula:

$$Y_{it}^{beh} = Y_{it}\mathbb{1}(Y_{it} > D_1) \tag{6}$$

Using this new outcome variable in model (DD) characterized above identifies the combined behavioral response (EXT+INT). In order to distinguish the EXT, INT, and AUT components, we further define the dependent variables Y_{it}^{ext}, Y_{it}^{int}, and Y_{it}^{aut} as follows:

$$\begin{aligned} Y_{it}^{ext} &= \mathbb{E}\left[Y_{it} \mid Y_{it} > D_1, X_{it}\right]\mathbb{1}(Y_{it} > D_1) \\ Y_{it}^{int} &= \left(Y_{it} - Y_{it}^{ext}\right)\mathbb{1}(Y_{it} > D_1) \\ Y_{it}^{aut} &= Y_{it}\left(1 - \mathbb{1}(Y_{it} > D_1)\right) \end{aligned} \tag{7}$$

Since the change in incentives is implemented at the insurance plan level, it is natural to cluster standard errors at the plan level and we do this in all specifications. However, given the relatively small number of clusters,[17] it is of interest to test the sensitivity of results to an alternative basis for inference. Thus, we also consider design-based inference, implemented in the following way: In a large number of iterations, 15 of the 25 plans in the analysis sample were randomly assigned to the treatment group and the rest to the control group. In each iteration, a placebo treatment effect was estimated and eventually the t-value of the main specification was compared to the distribution of t-values coming out of the placebo analysis (cf., MacKinnon and Webb, 2020). It is not possible to expose the entire range of outcomes we study to this permutation test: neither the automatic nor the intensive margin effects will be observable in the placebo draws. Therefore, we restrict ourselves to total claims and a dummy representing positive claims when conducting randomization inference.[18]

5. Data

We exploit rich claims data from a large private health insurer in Germany for the years 2005–2011 in our analyses. In order to study the effects of the no-claim refunds policy, we collapse information to the person-year-level and restrict our data in various ways. First, we only consider clients aged 25 or older in the base year, who have non-missing information — in order to rule out additional complications arising with insured children and families. Second, we restrict the sample to clients who remain with the insurer throughout the entire study period. This implies that we work with a balanced panel of clients.[19] Third, in our main analysis we only consider years 2005–2008 since no-claim refunds were subject to additional changes in subsequent years. However, in some specifications we extend our analysis until 2011 to study the dynamic impact of the refund policy while noting that additional revisions of the refund policy in later years may potentially obscure some of the results. Finally, all monetary variables (premiums, deductibles, claims, no-claim refunds) have been adjusted to account for inflation, and they are expressed in 2011 euros using the CPI provided by the German Federal Statistical Office.

[16] In an alternative specification, we also use models with individual fixed effects. See Table B.8 in the online Appendix for coefficient estimates for models with and without individual fixed effects.

[17] Since we also include clients who switch between plans outside the 2007–08 period in the analysis sample, the total number of clusters is larger than the 25 treated and control plans.

[18] We conduct a number of statistical tests, which might seem to justify adjustment for multiple testing. However in our case, total claims represents the primary endpoint, and all other outcomes we consider are derived from it. Hence, an established solution like e.g. the summary index suggested by Anderson (2008) would be equivalent to studying the estimated effect on total claims.

[19] In the short-term analysis for 2005–2008, the sample consists of clients who were enrolled each individual year, but we impose no requirement regarding the years 2009–2011. We relax the balanced-panel assumption as a robustness check, which yields similar point estimates when compared with the balanced sample.

Table 1
Overview of sampled insurance plans.

Group	Plans	No-claim size				Clients
		2005–2007	2008	2009	2010–2011	
Treatment	15	0.5/1/1.5/2	3	1/3	1/2/3	66,020 (53%)
Control	10	0.5/1	0.5/1	–	–	58,130 (47%)

Note.– Information on no-claim refunds is provided by the German private health insurer whose data is used for the analyses. Refund size measured in monthly premiums if claim-free for one/two/three/four consecutive years. Clients refer to the total number of individuals included in the main empirical analysis.

Table 1 provides an overview of the no-claim refund scheme for the insurance plans we include in our analysis sample. The intervention we study is a sudden increase in no-claim refunds for some insurance plans. These plans were offered by the health insurer throughout our analysis period. The change was announced in February 2008 and applied to claims made during the calendar year 2008. Hence, it is unlikely that clients were able to switch insurance plans because of the refund policy. For inclusion in the treatment and control groups, we require that clients remained within a treated or a control plan throughout the two-year period 2007–08; hence, individuals switching between these two groups, or between these groups of plans and other plans, are excluded from the analysis.

Our treatment group consists of more than 66,000 clients from 15 different plans for which the insurer changed the no-claim refund structure in 2008. Before the refund policy changed in 2008, an experience-rated refund was offered to all plans. Each additional year that a client did not submit a claim to the insurer was rewarded with a refund of 0.5 monthly premiums, up to a maximum of two monthly premiums after four claim-free years. In 2008, the no-claim refund structure for these plans changed and each client instead received a flat refund of three monthly premiums if no claims were filed in that year, regardless of their previous claiming history. As our comparison group, we select more than 58,000 clients from ten different plans with an unchanged refund structure from 2005 to 2008. Throughout the analysis period, these clients were subject 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.

Fig. 2 shows a graphical illustration of the potential no-claim refund size as a function of number of claim-free years by group to highlight the shifted monetary incentives from the refund policy. Prior to the introduction of the policy in 2008, both the treatment and control groups faced similar incentives to not submit claims to the insurer; the only difference was an extension of the scheme in the treatment group after the first two claim-free years. However, this option pertained to a very small group of contract holders.[20] In contrast, after the policy was introduced, members of the treatment group were provided with a large increase in the no-claim monetary incentive. Specifically, refund sizes climbed by up to 500 percent (from 0.5 to three monthly premiums) for clients with no claim-free years prior to the policy change. Since the potential refund size is directly linked to the client's total premium paid each year (e.g., three premiums corresponds to a total reduction in the annual premium of 25 percent), the incentives were likely to be both salient and financially relevant for a large group of clients.

We exclude the years after 2008 from our main analysis due to the additional changes in the refund structure that occurred in subsequent years. Specifically, Table 1 shows that the treatment group's refund structure featured a one-month premium payback for clients who had been claim-free for one year in 2009, and another two monthly premiums for clients who were claim-free for two years. This means that the overall generosity of the refund scheme was reduced in these years as the possibility to receive a maximum of three monthly premiums became subject to additional requirements. In addition, the refund option was altogether abolished in the control group from 2009 onward.[21] For this reason, we use the years 2009–2011 only to study long-run effects of the initial change in the refund policy, and we interpret these results with caution.[22]

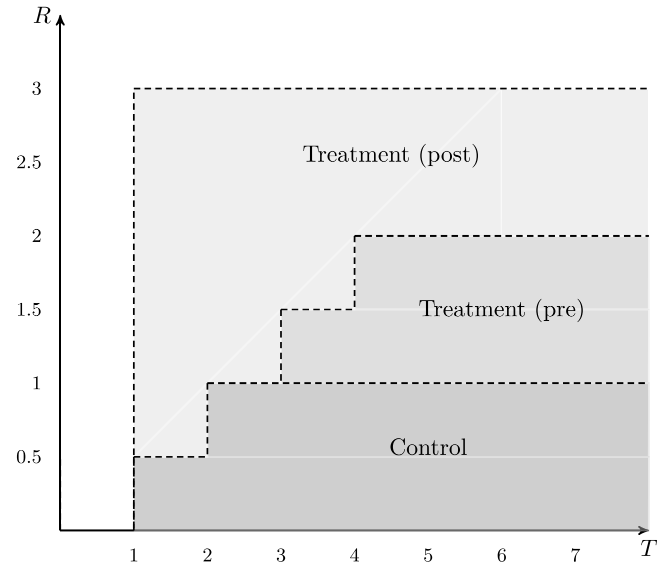


Fig. 2. Changes in potential No-Claim refund sizes from the 2008 policy change. Note.— No-claim refund sizes in months R as a function of claim-free years T by client group subject to the no-claim refund policy in 2008. Clients in the control group were unaffected by the policy.

Fig. 3 shows group-specific time trends in claims for clients in our analysis sample to assess the validity of our empirical approach. Panel (a) plots trends in total claims, while panels (b), (c) and (d) show trends in the three components from the decomposition method defined in Eq. (7). The parallel trend assumption, key for the causal interpretation of DID estimates, appears valid in all four cases.[23] Furthermore, the plots provide some evidence for a trend break for the treatment group in 2008; in particular for the probability of claiming a positive amount displayed in panel (b).

Table 2 reports summary statistics of variables other than claims in the sample for 2007, the year preceding the no-claim refund policy change.[24] We provide comparisons of demographics, occupational groups, insurance plan characteristics, and ACG scores. ACG scores are generated using the German version of the Johns Hopkins ACG software, which is routinely used by commercial insurers for underwriting

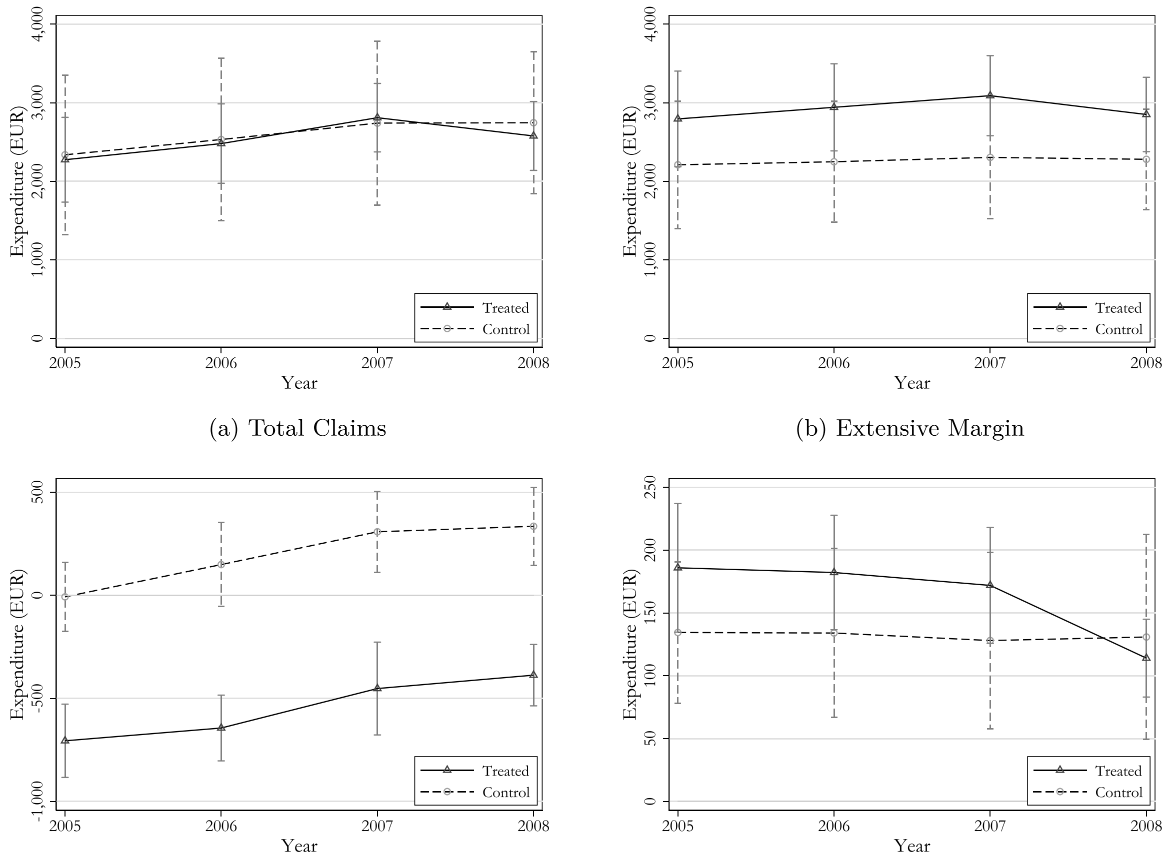
20 Figure B.4 in the online Appendix shows that only a small minority of clients were eligible for a refund of more than one monthly premium.

21 The subsequent changes in no-claim refunds in 2009 and thereafter were due to a deterioration of the insurer's financial situation as well as for the insurance sector as a whole (see discussion in online Appendix A).

22 In particular, we argue that a persistent long-run reduction in claims in the treatment group can be interpreted as a lower bound on the effect of the initial policy change in 2008, since the incentives to not claim in the treatment group were lower in 2009 and 2010. As this was also true for clients in the control group, who saw their refund option abolished altogether after 2008, we plot in Figure B.1 in the online Appendix trends in claims for both the treatment and control groups to show that our long-run effect estimates are not driven by changes in the control group.

23 The parallel trend assumptions for each group are also supported by formal statistical tests which are available upon request.

24 Table B.2 in the online Appendix reports summary sample statistics for all years. Specifically, the top and bottom panels of the table display statistics for the short (years 2005–2008) and the long (years 2005–2011) panel samples, respectively.



(a) Total Claims (b) Extensive Margin

(c) Intensive Margin (d) Automatic Effect

Fig. 3. Common trend graphs for claims: Short sample. Note.— Own calculations based on insurance claims data for years 2005–2008 described in Section 5. The analysis sample includes all claims for all clients aged 25 and above in 2005 who were enrolled during the entire analysis period. Panel (a) shows the total value of all submitted claims per person and year while panels (b)–(d) pertain to the resulting components from the decomposition method defined in Eq. (7) and derived in Section 4. The treatment group consists of insurance plans whose no-claim refund policy changed in 2008 and the control group consists of a set of plans that did not change the refund policy in this year.

Table 2
Descriptive statistics by treatment group 2007.

	Treated			Control			Comparison	
	N	Mean	SD	N	Mean	SD	Δ_{12}	Std. Dif
Female	66,020	0.114	0.318	58,130	0.285	0.452	−0.172	−0.440
Age	66,020	42.230	8.436	58,130	46.904	10.418	−4.675	−0.493
Employee	66,020	0.633	0.482	58,130	0.260	0.438	0.374	0.812
Self-Employed	66,020	0.334	0.472	58,130	0.669	0.470	−0.335	−0.712
Total claims	66,020	2.809	10.323	58,130	2.740	7.532	0.069	0.008
ACG Score	66,020	1.241	1.743	58,130	1.259	1.958	−0.018	−0.010
Risk premium	66,020	0.329	0.470	58,130	0.349	0.477	−0.020	−0.042
Exemptions	66,020	0.017	0.128	58,130	0.016	0.127	0.000	0.004
Deductible	66,020	0.200	0.153	58,130	1.011	0.188	−0.811	−4.733
Plus plan	66,020	0.215	0.411	58,130	0.341	0.474	−0.127	−0.285
Top plan	66,020	0.562	0.496	58,130	0.440	0.496	0.122	0.246

Note.– Own calculations based on insurance claims data for years 2005–2008 described in Section 5. The analysis sample includes claims for all clients aged 25 and above in 2005 who were enrolled in one of the insurance plans described in Table 1. The treatment group consists of insurance plans whose no-claim refund policy changed in 2008 and the control group consists of a set of plans that did not change the refund policy in this year. All monetary variables measured in thousands of 2011 euros. Standardized difference calculated according to Imbens and Wooldridge (2009).

purposes. The ACG software provides a continuous *risk score*, which represents the expected expenditure in a certain year, based on claims and on demographics in the previous year. We estimate ACG scores for each client in our data based on the combined pre-policy years in 2005–2007 in order to not confound this estimate with the effects of the refund policy change. An ACG score of one corresponds to mean expenditure in the German population; our sample has slightly higher scores, reflecting an over-representation of older individuals.

Clients in the control group are more likely to be female, older, and to have a higher deductible and a lower insurance premium. Furthermore, the control group is mostly represented by self-employed and the majority of clients in the treatment group are white-collar workers. On the other hand, ACG scores hardly differ between the two groups. While there are differences across treatment and control group with respect to clients' characteristics and claiming behavior, our empirical approach will produce valid causal estimates as long as

Table 3
Placebo difference-in-differences estimates.

	Mean	(1)	(2)	(3)	(4)	(5)	(6)
Total claims	2588	25.366 (115.989)	74.020 (90.867)	78.839 (90.552)	76.546 (90.628)	76.540 (90.651)	56.018 (87.275)
Positive claims	0.540	−0.007 (0.014)	−0.001 (0.009)	−0.001 (0.008)	−0.001 (0.008)	−0.001 (0.008)	−0.001 (0.006)
Very high claims	0.069	0.0014 (0.004)	0.0029 (0.004)	0.0032 (0.004)	0.0031 (0.004)	0.0031 (0.004)	0.0021 (0.004)
N		372,450	372,429	372,429	372,429	372,429	372,429
Year FE		✓	✓	✓	✓	✓	✓
Plan type FE		✓	✓	✓	✓	✓	✓
ACG Ventile FE			✓	✓	✓	✓	✓
Age & sex FE				✓	✓	✓	✓
Occupation FE					✓	✓	✓
Risk Class FE						✓	✓
Interactions							✓

Note.– Own calculations based on insurance claims data for years 2005–2007 described in Section 5. The analysis sample includes claims for all clients aged 25 and above in 2005 who were enrolled in one of the insurance plans described in Table 1. Each cell reports results from a separate regression. Reported coefficients pertain to γ^q estimates from estimation of Eq. (DD) for different outcomes (rows) and sets of regressors (columns). Estimates refer to outcomes measured in 2007, the year *prior* to the change in refund policy for the treatment group. All monetary variables measured in 2011 euros.

these differences are only manifested through varying outcome levels and the common trend assumption is not violated. To evaluate the robustness of our estimated results to confounding bias, we include flexible, non-parametric, specifications and group-specific interactions of all variables displayed in Table 2 in our empirical model.

6. Results

6.1. How did the no-claim refund policy affect claiming behavior?

We start our empirical investigation by estimating effects of the 2008 no-claim refund policy on total claims using our full analysis sample. The key identifying assumption for the estimation of causal effects in our application is that clients in treated and control plans would have followed a common time trend in their utilization behavior in absence of the policy. In order to find an appropriate empirical specification, we first conduct a placebo analysis by incorrectly assuming that the policy changed in year 2007. Based on an analysis sample covering the years 2005–2007, we estimate placebo "treatment effects" measuring deviations of the treatment group in 2007, the last pre-treatment year. As placebo outcomes we consider total claims, having positive claims, and having claims in the top ventile of the total claims distribution. Results are presented in Table 3, where each column reports point estimates from estimation of Eq. (DD) with sequential inclusion of control variables.

As can be seen from the table, point estimates are generally small and statistically indistinguishable from zero for all specifications. For example, the estimated effect on total claims is between one and three percent, depending on specification, and far from conventional levels of statistical significance (p-values range between 0.39 and 0.83). Estimation results for the other two outcomes convey similar messages. The inclusion of additional regressors has little impact on the placebo estimates, regardless of the baseline group differences in these characteristics. We nevertheless include these characteristics in our main DD model specification to improve precision.

Regression results using the same model and outcomes as in Table 3 for the reform year, 2008, are presented in Table 4. The point estimates for total claims are now highly significant across the board. The low variation in the coefficient estimates across specifications reinforces the conclusion that the control variables do not play an important role in predicting the effect of the refund policy. They do, however, substantially improve the precision of the estimates. Inspecting the magnitude of the estimated parameters in the first row, we see that clients exposed to the more generous no-claim incentives reduced their total claims by, on average, about €200 or eight percent of the average amount claimed in 2007. This result is statistically significant at the one percent level for our preferred specification in column (6) where the full set of control variables and interactions are included.[25]

The reported coefficients indicate that clients in the treatment group reduced their likelihood of claiming reimbursement for any medical expenses in 2008 by roughly ten percentage points. This is a precisely estimated, strongly statistically and economically significant effect, implying a 20 percent decrease in the propensity to claim compared to baseline (2007) levels. Intuitively, since the no-claim refund policy targeted the extensive margin of claiming, we expect the effect on overall claims to be mainly driven by claiming propensity. To further corroborate this hypothesis, we also estimate policy effects on the propensity to have very high claims as a falsification test. We do not expect high claimers to be affected by the policy, both due to their lack of control of their medical expenses (e.g., for individuals with severe chronic diseases) and the lack of incentives to respond (the refund is only a small fraction of the total expected medical expenditures for this group). Thus, the finding that the likelihood of very high claims is unaffected by the policy is reassuring and in line with our predictions.[26]

Table 5 presents the estimated effects of the refund policy by effect margin corresponding to the components in the decomposition method described in Section 4.[27] The first row of the table reports estimates for the overall behavioral effect, which is the sum of the extensive and intensive margin effects reported in rows two and three, respectively. The fourth row reports the non-behavioral, automatic effect, which is

[25] Results on the effects of patient cost-sharing schemes are often presented in terms of elasticities (see, e.g., Manning et al., 1987). However, Aron-Dine et al. (2015) point out that summarizing nonlinear contracts in the presence of dynamic incentives with a single price could be highly restrictive. Furthermore, we observe only claims, not total healthcare spending. For these reasons, we abstain from presenting our results in terms of elasticities.

[26] Table B.8 in the online Appendix reports results corresponding to the same specifications as in Table 4, but with individual fixed effects. The results are very similar.

[27] The distribution of healthcare expenditure is typically very skewed, and this goes for the data set we use here as well (cf., Karlsson et al., 2016). A consequence of the long right tail of the distribution is that the assumptions required for OLS estimation may be violated. In order to check whether this represents an issue in our case, we consider an alternative specification where we winsorize expenditure at the 99th percentile. This winsorizing is carried out for each year separately, so that we take the trends in expenditure into account. Table B.9 in the online Appendix provide the main results for the winsorized expenditure variable.

Table 4
Main difference-in-differences estimates.

	Mean	(1)	(2)	(3)	(4)	(5)	(6)
Total claims	2588	−228.500*	−200.779**	−192.329**	−194.050**	−193.897**	−196.508***
		(122.086)	(91.212)	(92.086)	(92.466)	(92.273)	(68.814)
Positive claims	0.540	−0.102***	−0.099***	−0.099***	−0.099***	−0.099***	−0.096***
		(0.017)	(0.013)	(0.013)	(0.012)	(0.012)	(0.007)
Very high claims	0.070	−0.0004	0.0004	0.0010	0.0009	0.0009	0.0007
		(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
N		496,300	496,272	496,272	496,272	496,272	496,272
Year FE		✓	✓	✓	✓	✓	✓
Plan type FE		✓	✓	✓	✓	✓	✓
ACG Ventile FE			✓	✓	✓	✓	✓
Age & sex FE				✓	✓	✓	✓
Occupation FE					✓	✓	✓
Risk Class FE						✓	✓
Interactions							✓

Note.– Own calculations based on insurance claims data for years 2005–2008 described in Section 5. The analysis sample includes claims for all clients aged 25 and above who were enrolled in one of the insurance plans described in Table 1. Each cell reports results from a separate regression. Reported coefficients pertain to γ^q estimates from estimation of Eq. (DD) for different outcomes (rows) and sets of regressors (columns). Estimates refer to outcomes measured in 2008, the year of the change in refund policy for the treatment group. All monetary variables measured in 2011 euros.

Table 5
Difference-in-differences estimates by effect margin.

	Mean	(1)	(2)	(3)	(4)	(5)	(6)
Behavioral effect	2028	−157.930	−137.424	−118.915	−120.835	−121.177	−120.469
		(126.303)	(104.970)	(108.475)	(108.909)	(108.990)	(78.765)
Extensive margin	2240	−185.605	−165.930*	−168.096	−168.469	−167.938	−135.519***
		(115.035)	(97.541)	(102.994)	(103.145)	(102.765)	(44.638)
Intensive margin	−213	27.675	28.505	49.181	47.633	46.761	15.049
		(47.311)	(44.486)	(44.182)	(43.404)	(43.501)	(42.648)
Automatic effect	560	−70.570***	−69.934***	−70.208***	−70.151***	−70.220***	−74.802***
		(11.619)	(11.458)	(11.233)	(11.237)	(11.221)	(10.154)
N		496,300	496,264	496,264	496,264	496,264	496,264
Year FE		✓	✓	✓	✓	✓	✓
Plan type FE		✓	✓	✓	✓	✓	✓
ACG Ventile FE			✓	✓	✓	✓	✓
Age & sex FE				✓	✓	✓	✓
Occupation FE					✓	✓	✓
Risk Class FE						✓	✓
Interactions							✓

Note.– Own calculations based on insurance claims data for years 2005–2008 described in Section 5. The analysis sample includes claims for all clients aged 25 and above in 2005 who were enrolled in one of the insurance plans described in Table 1. Each cell reports results from a separate regression. Reported coefficients pertain to γ^q estimates from estimation of Eq. (DD) for different outcomes (rows) and sets of regressors (columns). Outcomes refer to total claims and its resulting components from the decomposition method defined in Eq. (7) and derived in Section 4. Estimates refer to outcomes measured in 2008, the year of the change in refund policy for the treatment group. All monetary variables measured in 2011 euros.

the mechanical impact of a change in the claiming threshold, from D_0 to D_1, due to the refund policy. Comparing the first and fourth rows with the estimated effect on total claims from Table 4, we see that the behavioral effect accounts for approximately two-thirds of the total effect and the automatic effect for the remaining third. However, while the automatic effect is statistically significant at the one percent level, the behavioral effect is not. This curious result is explained when separately inspecting the extensive and intensive effects. In line with our predictions and previous findings from Table 4, the behavioral effect is entirely driven by the extensive margin effect, which is statistically significant at the one percent level for our preferred specification in column (6) of the table. In contrast, the intensive margin effect is estimated with positive sign. This is consistent with the second hypothesis from Section 3. However, the estimate for this margin is not statistically significant and does not rule out reductions also on the intensive margin. Thus, the combined behavioral effect is statistically insignificant due to the noise introduced by the intensive margin effect.

Finally, longer-term effects for years 2009–2011 are presented and compared with the short term effects for 2008 in Table 6. Here we only present separate regression estimates for total claims and the behavioral part of the effect (i.e., ignoring changes in the claims distribution below D_1). If clients are rational, we expect to see increased claims for the treatment group in subsequent years due to the refund scheme becoming less generous (or if clients deferred essential healthcare to obtain the refund). As can be seen from the table, the reported point estimates reject this hypothesis since the longer-term reduction in claims is even larger than the short-term effect.[28] This result bears a resemblance with Alalouf et al. (2019), who find that spending patterns are changed persistently even after the initial shock, in their case a diagnosis, has vanished. One potential explanation for this curious result is that clients became used to the idea of receiving a refund after 2008 and attempted to obtain it each year, which may have led them to further cut back on healthcare services.

[28] Since no-claim refunds were abolished for the control group in the year 2009, this effect can be interpreted as a lower bound. Figure B.1 in the online Appendix plots event studies of the effect margins for the long sample and by group. The estimated effect pattern suggests that the long-term effect of the policy is not caused by changes in the control group.

Table 6
Long-term effects.

	Total claims		Behavioral part	
Effect 2008	−244.7095**	−203.1121**	−170.5115	−127.1999
	(122.984)	(86.991)	(127.232)	(96.213)
Long-Term 2009–11	−332.8479***	−276.6047***	−250.7101**	−191.0091***
	(116.998)	(61.017)	(117.160)	(63.640)
Baseline	2588	2588	2028	2028
N	751,039	750,969	751,039	750,969
Year FE	✓	✓	✓	✓
Plan Type FE	✓	✓	✓	✓
ACG Ventile FE		✓		✓
Age & sex FE		✓		✓
Occupation FE		✓		✓
Risk Class FE		✓		✓
Interactions		✓		✓

Note.– Own calculations based on insurance claims data for years 2005–2011 described in Section 5. The analysis sample includes claims for all clients aged 25 and above in 2005 who were enrolled in one of the insurance plans described in Table 1. Each column reports results from a separate regression. Reported coefficients pertain to γ^q estimates from estimation of Eq. (DD) for different samples, outcomes, and sets of regressors indicated in the table. Outcomes refer to total claims and the behavioral component from the decomposition method defined in Eq. (6) and derived in Section 4. Estimates refer to outcomes measured in alternatively the year 2008, and in the years 2009–2011. All monetary variables measured in 2011 euros.

6.2. Who reacted to the no-claim refund policy?

Our main estimates of the effect of the no-claim refund policy showed that clients exposed to the policy on average reduced their claims by about eight percent. In this section we shed further light on this finding by studying in more detail which clients reacted to the policy in a series of heterogeneity analyses. We first report results by health status and subsequently results by changes in financial incentives.

6.2.1. Heterogeneity by health status

Effects by ACG score

The refund is conditional on the client not submitting any claims during the entire calendar year. In order to benefit from this policy, the client must reduce their healthcare expenditures to a level that provides a net financial surplus from receiving the refund and paying remaining expenses out-of-pocket. This condition requires that patients are relatively healthy, and it is unlikely to be relevant for most insurance holders with large annual medical expenses, such as those with severe chronic conditions. Therefore, we predict that the policy response should vary by the healthcare costs an individual client is expected to incur in a given year.

To investigate effect heterogeneity by individual health status, we first estimate a measure of each client's expected consumption of healthcare services using the Johns Hopkins Adjusted Clinical Group (ACG) system and all medical diagnoses recorded for the individual in the pre-treatment years. We then use this measure to estimate our DID model conditional on clients' relative position in the ACG score distribution. Fig. 4 illustrates the effect by ventiles of the ACG score distribution (bars indicate baseline spending by ventile) and the effect from the refund policy on total healthcare spending in 2008.[29] Gray and red lines in the figure refer to total and behavioral effect sizes as previously defined. Interestingly, when allowing the effect to vary in this manner, we see that the bulk of the response is concentrated among clients with ACG scores between ventiles five and nine. Furthermore, this is also the area of the ACG score distribution where point estimates are statistically significant at the five percent level (as indicated by the dashed lines representing confidence intervals). In contrast, the policy did not significantly impact claiming among clients belonging to ACG ventiles below five and above nine.

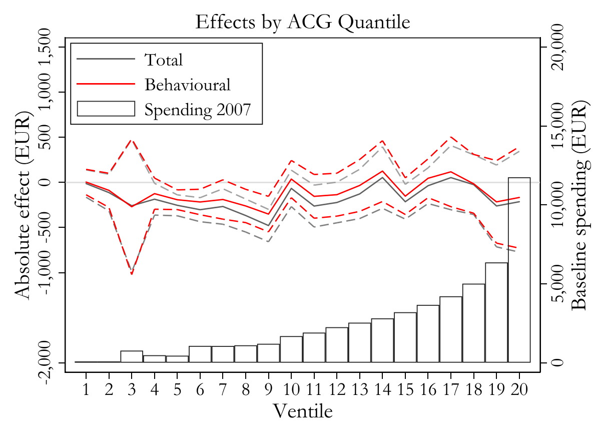


Fig. 4. Effect sizes by ACG ventile. Note.— Own calculations based on insurance claims data for years 2005–2008 described in Section 5. The analysis sample includes claims for all clients aged 25 and above in 2005 who were enrolled in one of the insurance plans described in Table 1. Line plots refer to estimated effects on total claims with associated 95 percent confidence intervals by ACG ventile, using the Johns Hopkins Adjusted Clinical Group (ACG) system and all medical diagnoses that have been recorded for the individual in 2005–07. Bars refer to ventile-specific average claims in 2007. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Do estimates reflect needs or preferences for healthcare?

Even though the ACG system has been designed to extract a health proxy from claims data, there is a potential concern that the algorithm picks up other individual attributes that are relevant in the response to the change in refund policy. This holds in particular for individual preferences for healthcare: a client in a low ACG ventile may either be healthier than the overall population, or have a weaker preference for healthcare consumption, or both. In order to investigate this issue further, we exploit a variable in the data set that can be argued to mainly reflect preferences: the plan generosity indicator which divides plans into three categories denoted ECO, PLUS and TOP.[30] The decision to join a certain type of plan will to some extent be guided by individual

[29] Table B.4 in the online Appendix reports point estimates of the coefficients and standard errors for each ventile by effect margin.

[30] ECO plans lack coverage for services such as single rooms in hospitals and treatments by a leading senior M.D. (*Chefarztbehandlung*) that TOP and PLUS plans offer. For ECO and PLUS plans, a 20 percent coinsurance rate applies if enrollees see a specialist without referral from their primary care physician, while such coinsurance does not apply for TOP plans.

health. Unsurprisingly, there is a strong gradient in the proportion of clients in generous plans along the ACG score distribution.[31] However, we posit that, *conditional* on baseline ACG scores, the plan generosity category should mainly reflect preferences for healthcare and, *conditional* on plan category, ACG scores will mainly reflect differences in healthcare needs.

We exploit this idea in two different ways. First, we reestimate Fig. 4 using two alternative specifications. In the first specification, we apply inverse probability weights to make the plan category indicator orthogonal to the ACG ventiles. Hence, we not only control for plan generosity as in our main specification, but also ensure that every ACG ventile gets an identical distribution across plan categories.[32] This strategy is thus reminiscent of the doubly robust DID estimators proposed by Sant'Anna and Zhao (2020). We contrast this using a second specification in which we do not control for plan categories at all. If estimates differ widely across these two model specifications, we interpret this as evidence that preferences, rather than need, for healthcare mediate the results reported in Fig. 4. The estimates in these two "extreme" specifications are remarkably similar and neither would change any of the conclusions drawn earlier.[33] We therefore conclude that preferences appear to be of secondary importance for the results represented there.

In order to shed further light on the role of preferences for healthcare, we also study effect heterogeneity for the three plan categories.[34] The top panel of Table 7 reports unweighted point estimates by plan category. The strongest response to the policy is noted for the middle (PLUS) category: clients in this group reduce their claims by €282 on average, of which €239 can be attributed to behavioral mechanisms. The least generous ECO plans have similar but lower reductions, whereas the most generous TOP plans have no reduction in claims at all. On the other hand, Column (5) reports that the TOP plans have the largest increase in the probability of claiming a refund. This change is likely driven by the automatic effect, which reflects that a large part of the density of the claims distribution for this group is located between the old and the new co-insurance thresholds.

It is possible that the effect heterogeneity reported in Table 7 reflects TOP clients being in worse health than the rest. In order to investigate how much of the difference can be attributed to differences in health, we present results from an entropy balancing specification in the lower part of the table. We implement the approach suggested by Hainmueller (2012), targeting the first, second and third moments of the pre-treatment realizations of ACG scores, age and sex.[35] The bottom panel of Table 7 shows the effect heterogeneity results for the entropy-balanced specification. The results remain similar to the unweighted estimates in all relevant aspects.

One peculiar feature of the results in Table 7 is that the TOP category has a significant increase in the probability of very large claims. This raises the question as to whether there are negative and positive effects of this group that cancel out. In order to shed some light on this issue, we estimate the impact of the treatment along the entire distribution of claims.[36] It turns out that the three groups of clients exhibit very similar reactions to the treatment for moderately-sized claims: claims below D_1 are reduced by similar amounts in all three groups. However, whereas holders of ECO and PLUS plans also reduce their probability of having positive claims in other parts of the distribution, the clients on TOP plans actually *increase* their probability of submitting claims slightly above D_1.

In sum, we find that there is substantial effect heterogeneity with respect to previous health in that individuals in ventiles five to nine of pre-treatment ACG scores were responsible for the bulk of the reduction in claims. This effect heterogeneity is robust to different methods of controlling for client preferences, as measured by their plan generosity. We also note significant effect heterogeneity between plan types. This heterogeneity is not affected by controlling for individual health. However, a closer examination reveals that all three groups respond as expected to the policy and in roughly similar amounts, and the effect heterogeneity is to a great extent driven by clients in the most generous plans increasing their utilization in parts of the distribution that should be unaffected by the policy. We therefore conclude that the main cause for effect heterogeneity are differences in healthcare needs. This result makes intuitive sense in terms of our theoretical framework since (i) very healthy clients were already unlikely to claim anything irrespective of the change in the refund policy and (ii) very unhealthy clients were unable to reduce their medical expenditures to a level that would produce a net financial gain.

Event study estimates

To study effect dynamics by incentive groups, Fig. 5 displays event study graphs of our main effects grouped into total claimed amount and its three components based on our decomposition method. The plot markers indicate year-by-year point estimates of the refund policy effect from years 2005 to 2011 for ACG score ventiles 5–9 (in red) and for all other categories (in gray), respectively. The effect pattern in each of the plots reinforce the interpretation from our previous findings: the overall effect is mainly driven by the extensive margin and automatic effects and by clients who had the most to gain from the policy by becoming eligible for the refund. Furthermore, in line with the results from Table 6, we see that the effects in subsequent years show no sign to reverse or even level off over time.

6.2.2. Heterogeneity by variation in financial incentives

Number of claim-free years

Next, we study how clients responded to changes in financial incentives from the policy in terms of previous claim-free years. According to the number of claim-free years prior to the refund policy change, some clients had more than others to gain from withholding claims. Fig. 2 provided a graphical illustration of the incentives produced by the 2008 policy change by the number of previous claim-free years. The change in the refund policy meant that the option value from previous claim-free years was lost due to the flat refund structure introduced by the 2008 policy. As can be seen from the figure, clients in the treatment group with four or more previous claim-free years would only gain one additional month's worth of refunds, while clients with no previous claim free years would gain 2.5 additional months. Although the nominal refund size after the increase in refunds is the same for all members of the treatment group, it is interesting to study whether a larger change in incentives leads to a larger effect of the policy reform.

We investigate this by using our baseline model to estimate heterogeneous effects for clients who received a refund in the previous year compared with all other clients. We consider two different specifications: first, we allow for heterogeneity by 2007 refund status, and

[31] See Figure B.5 in the online Appendix.

[32] We allow the weights to be different in the treatment and control groups. The weight for an individual in treatment group $d \in \{0,1\}$ belonging to ACG ventile a and plan category $j \in \{ECO, PLUS, TOP\}$ was calculated as $w_d^{aj} = N_d^j / 20 \cdot N_d^{aj}$ where N_d^j denotes the number of individuals who belong to plan category j in 2007, and N_d^{aj} denotes the number of individuals who belong to plan category j and ACG ventile a.

[33] The two corresponding figures are exhibited in Figure B.6 in the online Appendix.

[34] Table B.5 in the online Appendix provides descriptive statistics for the three subsamples which confirm that the more generous plans generate larger claims in general and therefore also require higher premiums.

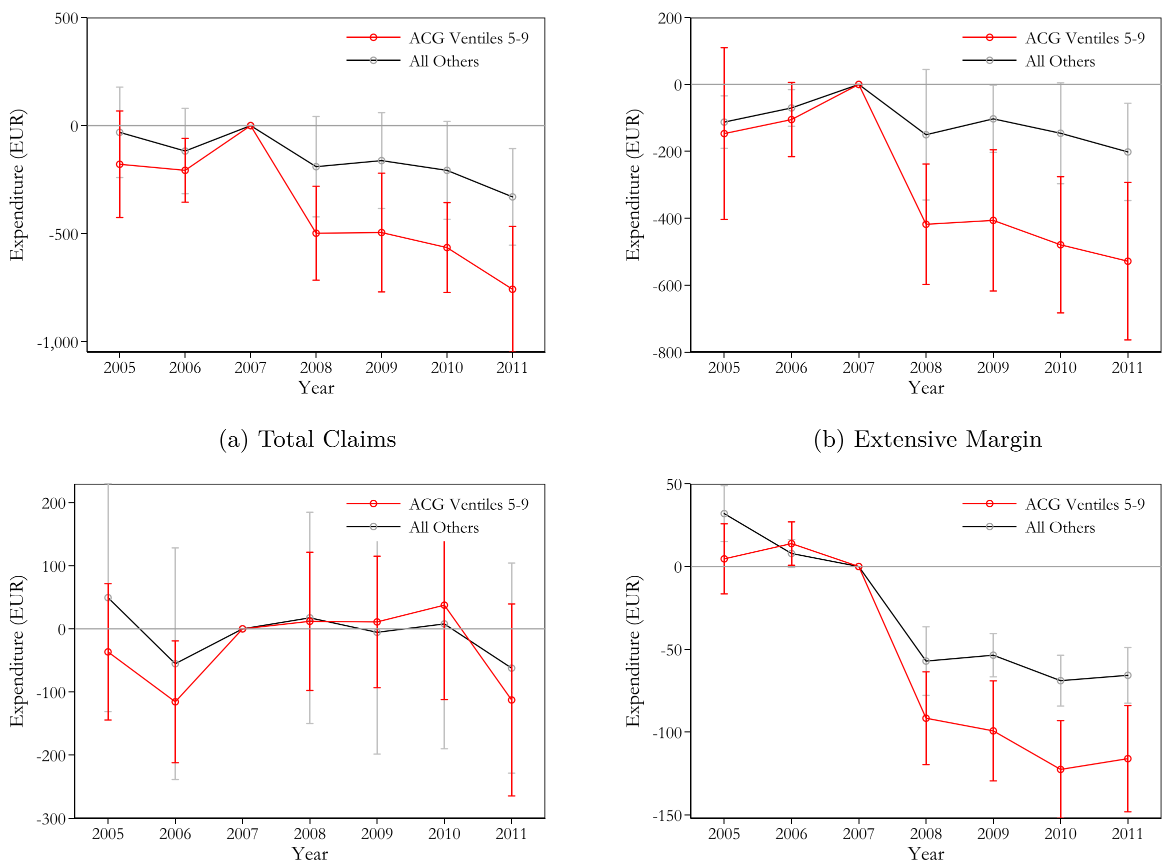
[35] The right panel of Table B.5 in the online Appendix displays the weighted means of certain variables and shows that the balancing works very well for all three variables but that some differences in claiming behavior remain even after removing the influence of these variables.

[36] Results are presented in Figure B.7 in the online Appendix.

Table 7
Results by plan category.

	Claims					Refund
	Baseline	Total	Behavioral	Automatic	Very high	
		(1)	(2)	(3)	(4)	(5)
Unweighted						
Cat ECO	1535	−179.085	−161.252	−17.833***	−0.004	0.067***
		(137.86)	(141.21)	(3.69)	(0.01)	(0.01)
Cat PLUS	1935	−281.924**	−238.870**	−43.054***	−0.004	0.096***
		(106.66)	(104.87)	(5.84)	(0.01)	(0.02)
Cat TOP	3248	−18.764	77.735	−96.499***	0.011***	0.146***
		(55.71)	(49.43)	(8.60)	(0.00)	(0.01)
Entropy balanced						
Cat ECO	1797	−187.126	−168.969	−18.157***	−0.004	0.064***
		(178.54)	(182.51)	(4.31)	(0.01)	(0.01)
Cat PLUS	1935	−278.944**	−235.835**	−43.109***	−0.004	0.096***
		(110.61)	(108.75)	(5.71)	(0.01)	(0.02)
Cat TOP	2498	−70.075	30.731	−100.806***	0.008**	0.158***
		(48.88)	(42.05)	(7.53)	(0.00)	(0.01)

Note.– Own calculations based on insurance claims data for years 2005–2008 described in Section 5. The analysis sample includes claims for all clients aged 25 and above in 2005 who were enrolled in one of the insurance plans described in Table 1. The top panel shows unweighted estimates of means and standard deviations of variables broken down by clients' 2007 plan generosity rating. The bottom panel consists of weighted estimates based on entropy balancing to make the first, second and third moments of ACG scores, age, and sex match the distribution of the PLUS category (cf., Hainmueller and Xu, 2013). All monetary variables measured in 2011 euros.



(a) Total Claims

(b) Extensive Margin

(c) Intensive Margin

(d) Automatic Effect

Fig. 5. Event studies for claims by ACG category at baseline. Note.— Own calculations based on insurance claims data for years 2005–2011 described in Section 5. The analysis sample includes claims for all clients aged 25 and above in 2005 who were enrolled in one of the insurance plans described in Table 1 for the entire analysis period. Panel (a) shows the total value of all submitted claims while panels (b)–(d) pertain to the components from the decomposition method defined in Eq. (7) and derived in Section 4. The treatment group consists of insurance plans whose no-claim refund policy changed in 2008, and the control group consists of a set of plans that did not change the refund policy in this year. Red and gray lines refer to treatment effects for ACG ventiles 5–9 and all other ACG ventiles, using the Johns Hopkins Adjusted Clinical Group (ACG) system and all medical diagnoses that have been recorded for the individual in 2005–07. All monetary variables measured in 2011 euros. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

second, we allow for heterogeneity by refund status in the previous years. The two specifications have different implications for the time trend assumptions and require different sample selection rules, but essentially test the same thing: whether there is effect heterogeneity by the pre-treatment propensity to submit claims. Results from both specifications suggest that clients who had no refund in the previous year respond more than other clients.[37] This is consistent with their stronger incentives for reducing claims. On the other hand, this group also has, by construction, a higher probability of submitting claims and therefore also a greater possibility to change their behavior in response to these incentives.

Inframarginal non-users and loss of insurance value

Some clients who were already non-claimers before the refund policy was introduced received a windfall gain from the policy as they received a larger refund without actively changing their claiming behavior. This meant in practice that the policy transferred insurance funds from users to these "inframarginal" non-users. We study the characteristics of this group and compare them with those that did submit claim in one or several years during our analysis time frame. We find that inframarginal non-users are younger and healthier, less likely to be female or employed (in contrast to being self-employed), and more likely to be enrolled in a low-cost insurance plan.[38] These characteristics apply irrespective of whether we condition on the treatment group or not. Thus, the refunds policy indeed shifted resources from clients with higher demand, needs, and preferences for healthcare to clients with low or no demand, needs, or preferences.

The increase in no-claim refunds implied less transfers from people in healthy (low marginal utility) states to people in unhealthy (high marginal utility) states. To assess the corresponding loss of insurance value, we conduct a back-of-the-envelope calculation. We assume preferences with constant absolute risk aversion (CARA) and a standard value of risk aversion of 10^{-5}, a probability of receiving a no-claim refund of 0.5, and an additional no-claim refund of 2 monthly premiums of €250 each.[39] Then, the risk premium of the increased no-claim refund is €0.000625.[40] Even for a very high risk aversion of 10^{-3} the risk premium for the increased no-claim refund is only €0.0625. Hence, the loss of insurance value is small relative to the reduction in spending.

6.3. What claims did clients cut down on?

Next, we explore the types of healthcare that patients cut down on because of the refund policy. While our data do not allow us to distinguish utilization according to a high-value/low-value dichotomy (cf. Schwartz et al., 2014; Brot-Goldberg et al., 2017), we are able to single out some treatments that may be non-essential for reasons we specify further below. The results from this analysis will inform policy regarding the efficiency of no-claim refund arrangements in terms of reducing unnecessary care use.

6.3.1. Results by claim type

Results by claim type are presented in Table 8. We present the overall effect according to our preferred specification alongside breakdowns by service type, and we contrast the immediate effect (in 2008) with the long-term (2009–11) effect.[41] A first such breakdown that we consider is to use the diagnosis associated with each claim. We calculate the total amount claimed for specific diagnoses, and we contrast it with the total amount claimed without diagnosis information; the latter amount includes claims with missing information and claims that have been assigned ICD chapter R ("Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified"). Claims falling within this classification often encompass treatment episodes lacking medical justification. However, the absence of a diagnosis code may also stem from other factors, such as genuine uncertainty on the part of the healthcare provider. In addition, we provide breakdowns for three particular types of services. One peculiarity of the German health insurance system is that alternative medicine treatment services, for which the empirical evidence of effectiveness is either weak or nonexistent, are reimbursable. We posit that services in this broad category can be considered as low-value by definition, at least according to their economic value. Furthermore, we explore the impact on claims related to visual aids and dental care. These categories may constitute high-value care both from an individual and societal standpoint. However, they are often deferrable without causing immediate harm to one's health.[42]

Results in Table 8 show that unspecific claims are disproportionately responsible for the reduction in total claims: total unspecific claims are reduced by 9.9 percent, compared to 6.5 percent for specific claims. When we focus on the behavioral part of the change, the effect on unspecific claims is similarly considerably larger than the effect on specific claims. Even though the diagnostic information entailed in a claim is at best an imperfect proxy for low-value care, this finding suggests that clients consider care with unspecified diagnoses as lower value than care with specified diagnoses. Notably, we find the opposite result for alternative medicine. However, the estimated effect is insignificant and small in absolute and relative terms.

In terms of the two "deferrable" care categories we consider, we see that the reduction in dental care expenditure is lower than the overall reduction. On the other hand, there is a stronger reduction on spending on visual aids.[43] We also note that the initial impact of the policy remains persistent in most cases: the claim types that respond disproportionately in 2008 also contribute disproportionately to the long-term effect, and vice versa. The exception to this pattern is alternative medicine, which portrays a relatively stronger reduction over time compared to other expenditure categories.

In general, Table 8 delivers evidence supporting the view that the medical value of treatments was an important factor when clients reacted to the policy. However, the reaction pattern is more complex than a simple prioritization of high-value care at the expense of low-value care. This may to some extent be due to the clients' valuation of different types of care deviating from the social valuation. For example, the weak initial response in spending on alternative medicine may reflects users' high subjective valuation of such treatments.

6.3.2. Results by diagnosis group

The specified claims in our data are represented by diagnosis codes classified according to WHOs ICD-10 standard. We use this information to estimate diagnosis group-specific effects of the refund policy by ICD-10 chapter. Results from this analysis are shown in Fig. 6 for 2008 (left panel) and 2008–2011 (right panel). The vertical axis displays the effect size in relative terms (i.e., as a share of the baseline value in 2007), while the horizontal axis reports absolute effects in euros. Moreover, red colored markers highlight effects that are significant at the ten percent level, while gray markers indicate insignificant effects. As can be seen, most ICD chapters are associated with small and statistically insignificant point estimates for the short-run effect

[37] Results are reported in Table B.6 in the online Appendix for total claims and each of the claim components.

[38] Results are shown in Table B.3 in the online Appendix.

[39] We choose the value of risk aversion following Handel et al. (2020).

[40] Following Pratt (1964), we compute the risk premium as $0.5 \cdot \sigma^2 \cdot r = 0.5 \cdot 125 \cdot 10^{-5} = 0.000625$ where σ^2 is the variance of additional out-of-pocket payments, and r is the value of risk aversion.

[41] We show raw time trends for each component considered in this analysis in Figure B.2 in the online Appendix.

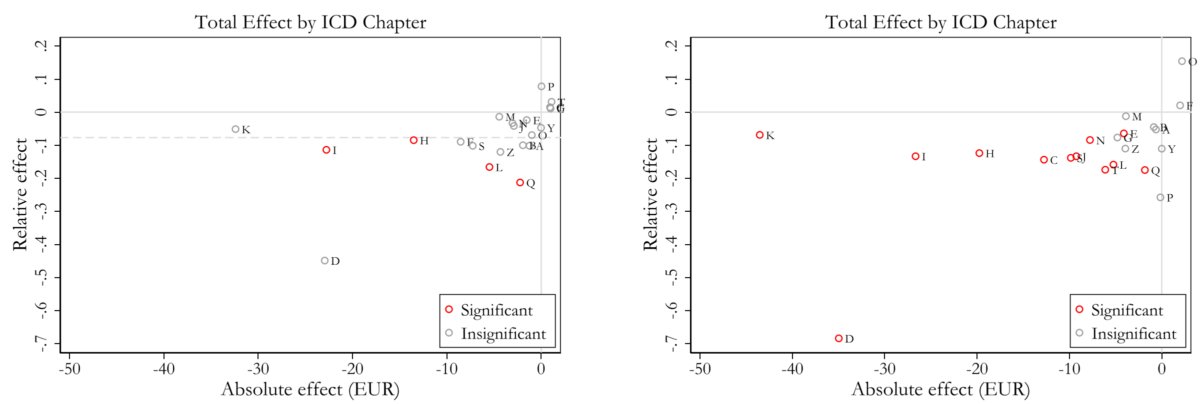
[42] Online Appendix B.3 reports evidence suggesting that this breakdown of claims reflects differences in clients' behavior: exploiting variation in claiming intensity within a year, we show that the propensity to consume these service types changes in different directions when the price drops to zero.

[43] The overall response for visual aids and dental care is in line with the within year change in demand according to the analysis in online Appendix B.3. Visual aids, for which demand increases once the effective price drops to zero, respond disproportionately. Dental care, which is less price sensitive, also contributes less to the change in no-claim refunds.

Table 8
Difference-in-differences estimates by claim type.

	Baseline	Effect 2008				Long-Term 2009-11			
		Total		Behavioral		Total		Behavioral	
		€	%	€	%	€	%	€	%
Total claims	2809	−207.5**	−7.4	−131.4	−5.0	−284.1***	−10.1	−198.0***	−7.5
		(86.2)		(95.2)		(58.9)		(61.4)	
Specific claims	2068.8	−134.9*	−6.5	−81.3	−4.2	−191.6***	−9.3	−133.3***	−6.8
		(81.0)		(85.8)		(46.4)		(48.5)	
Unspecific claims	746.8	−73.8***	−9.9	−48.4***	−7.0	−100.5***	−13.5	−64.6**	−9.3
		(13.6)		(14.2)		(37.1)		(30.6)	
Alternative medicine	45.4	−0.9	−1.9	0.4	1.0	−4.2*	−9.2	−2.3	−5.5
		(1.4)		(1.8)		(2.2)		(1.9)	
Visual aids	97.1	−11.2***	−11.6	−5.3	−6.6	−10.2***	−10.5	−1.3	−1.6
		(1.8)		(3.2)		(1.9)		(2.3)	
Dental care	569.5	−30.5*	−5.3	−14.9	−2.8	−48.2***	−8.5	−30.1*	−5.7
		(16.8)		(26.1)		(13.5)		(16.5)	

Note.– Own calculations based on insurance claims data for years 2005–2008 (left columns) and years 2005–2011 (right columns) described in Section 5. The analysis sample includes claims for all clients aged 25 and above in 2005 who were enrolled in one of the insurance plans described in Table 1. Each cell reports results from a separate regression. Reported coefficients pertain to γ^q estimates from estimation of Eq. (DD) for different outcomes and sets of regressors indicated in the table. Unspecific claims are claims without an ICD-10 code, or with an ICD-10 code in Chapter R ("Symptoms and signs not elsewhere classified"). Outcomes refer to total claims and the behavioral component from the decomposition method defined in Eq. (6) and derived in Section 4. All monetary variables measured in 2011 euros.



(a) Effect 2008 (b) Effect 2008-11

Fig. 6. Difference-in-Differences estimates by ICD-10 chapter. Note.— Own calculations based on insurance claims data for years 2005–2011 described in Section 5. The analysis sample includes claims for all clients aged 25 and above in 2005 who were enrolled in one of the insurance plans described in Table 1. Estimates refer to outcomes measured in the year of the change in refund policy for the treatment group (left panel) or over the pooled post-policy period 2008–2011 (right panel). Each marker reports results from a separate regression. Reported coefficients pertain to γ^q estimates from estimation of Eq. (DD) for different outcomes. Absolute effects (horizontal axis) refer to face value estimates of γ^q while relative effects (vertical axis) refer to effect sizes as a share of the total chapter-specific claims in 2007. Outcomes refer to specific ICD-10 chapters as indicated by marker labels. Claim diagnoses are classified according to the International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)-WHO Version 2016 (see https://icd.who.int/). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

outcome. The exceptions are for circulatory diseases (I), diseases of the eye and adnexa/ear and mastoid process (H), diseases of the skin and subcutaneous tissue (L) and congenital malformations, deformations and chromosomal abnormalities (Q).[44] It is worth noting that the effect size of the two latter categories, although statistically significant, are very close to zero and economically negligible. Conversely, the effects for diseases of the digestive system (K) and diseases of the blood and blood-forming organs (D) are large in absolute terms but imprecisely estimated.

To increase statistical precision, we instead focus on the estimates for the entire post-policy period between 2008 and 2011. Interestingly, the inclusion of the additional years retains the rank order of categories in terms of effect sizes. In addition, ICD-10 chapters K and D are now significant owing to the increased statistical precision from the larger sample.[45]

In summary, our findings from this subsection provide a mixed and intricate pattern of the types of care that clients primarily cut down on because of the refund policy. One the one hand, we observe a clear reduction in discretionary treatments, proxied by unspecified claims, typically initiated by patients and which might not have immediately observable negative health consequences if they are avoided or postponed. On the other hand, we find no indications that alternative medicine treatments, an indicator for low-value care, were significantly affected by the refund policy; nor do we find a clear pattern in the results for our selected indicators of deferrable care: dental care and

[44] We provide information on common components of these chapters in online Appendix B.

[45] However, results for chapter D are driven by a single observation in the year 2007. If we winsorize the outcome variable the absolute effect for chapter D is relatively small, and the relative effect is more comparable to other chapters. Results for other chapters are not driven by outliers. Figure B.9 in the online Appendix presents estimates by ICD chapter based on winsorized data.

visual aids. Since we are unable to follow clients beyond 2011 due to lack of data, we can only speculate about the potential long-term effects of the refund policy. However, the specific disease categories primarily affected by the no-claim incentive indicate that clients may have at least partially cut down on care which can be considered as high value, which is in line with previous findings by Manning et al. (1987) and Brot-Goldberg et al. (2017).

6.4. Robustness checks

Assumptions on Claiming Behavior

The interpretation that our estimates reflect demand for healthcare services requires that the "claiming function" presented in Eq. (3) represents a reasonable approximation of client behavior. This claiming function relies on strong assumptions, as outlined in Section 4.1. We assume that patients file claims if and only if their annual healthcare expenditures are above the claiming threshold in Eq. (3). This assumption could be violated if insurance holders submit claims below the claiming threshold. Such instances may arise due to irrational behavior, lack of awareness regarding the no-claim refund threshold, or liquidity constraints on the part of insurance holders. In Figure B.3 in the online Appendix we show that most insurance holders either submit no claim or claims above the threshold, but there are also some individuals who submit claims below the threshold. Table B.12 in the online Appendix shows that our main estimation results are robust if we exclude individuals from our sample who violated the postulated "claiming function" in pre-treatment years by submitting positive claims below the deductible threshold. This suggests that our results are not confounded by clients who act irrationally or who are liquidity constrained.

The claiming function could also be violated if patients with healthcare expenditures above the threshold do not submit claims, for example because of hassle costs. From the insurance company we know that online submission was not implemented during our study period. Accordingly, clients faced moderate hassle costs by collecting bills in paper format and submitting them via standard mail. However, this was a rather informal process and no official forms nor specific requirements were necessary. Yet, a higher no-claims refund does not affect hassle costs. If anything, it increases incentives to submit claims. Thus, our estimation results can be seen as a lower bound of the true effect.

Theoretically, clients might face a degree of uncertainty whether and to what extent their submitted claims will be reimbursed by the insurer and incorporate this uncertainty in their claiming behavior. However, we consider this to be of lesser importance in the German health insurance system. Inpatient services are typically reimbursed directly between hospitals and the insurer due to the high costs encountered here. For ambulatory costs exists a transparent system (the so-called Gebührenordnung für Ärzte, GOÄ) that clearly states services and prices that are uniform across providers for clients of private health insurers. Prices are uniform for all healthcare providers and private insurers, and the list of services covered is comprehensive. In Germany, there is no selective contracting, and there are no out-of-network providers. Accordingly, we assume that uncertainty with respect to reimbursement plays a negligible role in the decision of clients. We confirmed this in contact with experts of the insurance company.

Functional Form of Healthcare Claims

Healthcare expenditure is a canonical case of a variable exhibiting heavy tails (French and Jones, 2004; Karlsson et al., 2023). In order to assess the sensitivity of our results to heavy tails, we reran the analysis with expenditures winsorized at the 99th percentile within each year. Our main results for winsorized expenditure are provided in Table B.9 in the online Appendix. The results do not change qualitatively and the estimates are hardly affected in a quantitative sense. Hence, our main findings are not driven by outliers.

Furthermore, we estimate two-part models that account for the large number of zero claims in addition to the heavy right tails in the distribution of claims (Deb and Norton, 2018). Our specification tests suggest a Power link function and a Gaussian distribution of the error term (see Table B.10 in the online Appendix). We present results for this preferred specification as well as for a two-part model with a Log link function and a Gamma distribution of the error term in Table B.11 in the online Appendix. Effect sizes for the two-part models are somewhat larger, but overall similar in magnitude compared to our baseline linear models.

Statistical Inference

As an additional robustness check, we conduct the permutation test outlined in Section 4.3 (cf. MacKinnon and Webb, 2020). The results are presented in Figure B.10 in the online Appendix for total claims (left figure) and the probability of having positive claims (right figure). The estimate for total claims is significant at the 3.2 percent level, which is somewhat less precise compared to the statistical inference reported in Table 4 suggests. However, the estimate for positive claims is significant at any conventional level according to this alternative basis for inference.

7. Conclusion

No-claim refunds are cost-control instruments which stipulate a payback agreement contingent on a claim-free calendar year by the insured. We study how economic incentives derived from no-claim refunds affect claiming behavior using rich administrative claims data from a large German health insurer and an insurer policy that unexpectedly increased the refund size for certain insurance plans. Our results show that clients in plans that were subject to an increase in refunds significantly reduced their claims by eight percent (€200) on average, relative to clients with plans for which refunds were unchanged.

Using a novel method to decompose the overall effect into an intensive, extensive and an automatic component, our findings show that individuals reacted to the changed incentives by reducing their claims mainly on the extensive margin. This finding is in line with theoretical predictions since the policy gave clients stronger incentives not to submit any claims. We further show that these behavioral responses were stronger for clients more exposed to the incentives. In addition, we show that no-claim refunds lead to changes in behavior that are likely to last even after the original incentives have been withdrawn, and that reductions in claims were not confined to treatments of questionable medical value.

In summary, clients seem to respond to changes of the design of health insurance plans with respect to no-claim refunds. Individuals seem to understand the dynamic and non-linear design of insurance plans and generally react in line with general intuition and economic theory. However, it is less clear whether they understand the potential risks and health consequences of delaying or avoiding important elective care. A potentially fruitful avenue for future research could therefore be to analyze long-term health effects of individual clients exposed to different cost-containment incentives in relation to the specific healthcare services they cut down on.

Declaration of competing interest

None

Data availability

The data that has been used is confidential.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jpubeco.2023.105061.

References

Abaluck, Jason, Gruber, Jonathan, Swanson, Ashley, 2018. Prescription drug use under Medicare Part D: A linear model of nonlinear budget sets. J. Public Econom. 164, 106–138.

AGPHI, 2020. Financial Report for Private Healthcare Insurance 2019. Report, AGPHI, Cologne, Germany.

Alalouf, Mattan, Miller, Sarah, Wherry, Laura R., 2019. What Difference Does a Diagnosis Make? Evidence from Marginal Patients. Working Paper 26363, National Bureau of Economic Research.

Anderson, Michael L., 2008. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and early training projects. J. Amer. Statist. Assoc. 103 (484), 1481–1495.

Aron-Dine, Aviva, Einav, Liran, Finkelstein, Amy, Cullen, Mark, 2015. Moral hazard in health insurance: Do dynamic incentives matter? Rev. Econ. Stat. 97 (4), 725–741. http://dx.doi.org/10.1162/REST_a_00518.

Atal, Juan Pablo, Fang, Hanming, Karlsson, Martin, Ziebarth, Nicolas R., 2017. Exit, voice, or loyalty? An investigation into mandated portability of front-loaded private health plans. J. Risk Insurance 86 (3), 697–727.

Atal, Juan Pablo, Fang, Hanming, Karlsson, Martin, Ziebarth, Nicolas R., 2020. Long-Term Health Insurance: Theory Meets Evidence. Working Paper 26870, National Bureau of Economic Research.

Bhargava, Saurabh, Loewenstein, George, Sydnor, Justin, 2017. Choose to lose: Health plan choices from a menu with dominated option. Q. J. Econ. 132 (3), 1319–1372. http://dx.doi.org/10.1093/qje/qjx011.

Blümel, Miriam, Spranger, Anne, Achstetter, Katharina, Maresso, Anna, Busse, Reinhard, 2020. Germany: Health system review. Health Syst. Transit. 22 (6), 1–272.

Brot-Goldberg, Zarek C., Chandra, Amitabh, Handel, Benjamin R., Kolstad, Jonathan T., 2017. What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics. Q. J. Econ. 132 (3), 1261–1318. http://dx.doi.org/10.1093/qje/qjx013.

Cabral, Marika, 2017. Claim timing and ex post adverse selection. Rev. Econom. Stud. 84 (1), 1–44. http://dx.doi.org/10.1093/restud/rdw022.

Dalton, Christina M., 2014. Estimating demand elasticities using nonlinear pricing. Int. J. Ind. Organ. 37, 178–191.

Dalton, Christina M., Gowrisankaran, Gautam, Town, Robert J., 2019. Salience, myopia, and complex dynamic incentives: Evidence from Medicare Part D. Rev. Econom. Stud. 87 (2), 822–869. http://dx.doi.org/10.1093/restud/rdz023.

Deb, Partha, Norton, Edward C., 2018. Modeling health care expenditures and use. Annu. Rev. Public Health 39, 489–505.

Einav, Liran, Finkelstein, Amy, Schrimpf, Paul, 2015. The response of drug expenditure to nonlinear contract design: Evidence from Medicare Part D. Q. J. Econ. 130 (2), 841–899. http://dx.doi.org/10.1093/qje/qjv005.

French, Eric, Jones, John Bailey, 2004. On the distribution and dynamics of health care costs. J. Appl. Econometrics 19 (6), 705–721. http://dx.doi.org/10.1002/jae.790.

Hainmueller, Jens, 2012. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. Polit. Anal. 20 (1), 25–46.

Hainmueller, Jens, Xu, Yiqing, 2013. Ebalance: A stata package for entropy balancing. J. Stat. Softw. 54 (7).

Handel, Benjamin R., Kolstad, Jonathan T, Minten, Thomas, Spinnewijn, Johannes, 2020. The Social Determinants of Choice Quality: Evidence from Health Insurance in the Netherlands. Working Paper 27785, National Bureau of Economic Research.

Hayen, Arthur P., Klein, Tobias J., Salm, Martin, 2021. Does the framing of patient cost-sharing incentives matter? The effects of deductibles vs. no-claim refunds. J. Health Econom. 80, 102520. http://dx.doi.org/10.1016/j.jhealeco.2021.102520.

Imbens, Guido W., Wooldridge, Jeffrey M., 2009. Recent developments in the econometrics of program evaluation. J. Econ. Lit. 47 (1), 5–86.

Johnson, Eric J, Hershey, John, Meszaros, Jacqueline, Kunreuther, Howard, 1993. Framing, probability distortions, and insurance decisions. J. Risk Uncertain. 7, 35–51.

Jürges, Hendrik, 2009. Health insurance status and physician behavior in Germany. J. Contextual Econom. 129 (2), 297.

Kahneman, Daniel, Tversky, Amos, 1979. Prospect theory: An analysis of decision under risk. Econometrica 47 (2), 263–291.

Kairies-Schwarz, Nadja, Rieger-Fels, Markus, Waibel, Christian, 2023. Cost-sharing or rebate: The impact of health insurance design on reducing inefficient care. Front. Behav. Econom. 2, 1.

Karlsson, Martin, Klein, Tobias, Ziebarth, Nicolas R., 2016. Skewed, persistent and high before death: Medical spending in Germany. Fiscal Stud. 37 (3–4), 527–559.

Karlsson, Martin, Wang, Yulong, Ziebarth, Nicolas R., 2023. Getting the Right Tail Right: Modeling Tails of Health Expenditure Distributions. Working Paper 31444, National Bureau of Economic Research.

Klein, Tobias J., Salm, Martin, Upadhyay, Suraj, 2022. The response to dynamic incentives in insurance contracts with a deductible: Evidence from a differences-in-regression-discontinuities design. J. Public Econom. 210, 104660. http://dx.doi.org/10.1016/j.jpubeco.2022.104660.

MacKinnon, James G., Webb, Matthew D., 2020. Randomization inference for difference-in-differences with few treated clusters. J. Econometrics 218 (2), 435–450.

Manning, Willard G., Newhouse, Joseph P., Duan, Naihua, Keeler, Emmett B., Leibowitz, Arleen, 1987. Health insurance and the demand for medical care: Evidence from a randomized experiment. Amer. Econ. Rev. 77 (3), 251–277.

OECD, Indicators, 2015. Health at a Glance 2011, Vol. 15. OECD Indicators, OECD Publishing, Paris, p. 2016. http://dx.doi.org/10.1787/health_glance-2015-en, Accessed February 2023.

Pauly, Mark V., 1968. The economics of moral hazard. Amer. Econ. Rev. 58 (3), 531–537.

Pratt, John W., 1964. Risk aversion in the small and in the large. Econometrica 32 (1/2), 122–136.

Remmerswaal, Minke, Boone, Jan, Bijlsma, Michiel, Douven, Rudy, 2019. Cost-sharing design matters: A comparison of the rebate and deductible in healthcare. J. Public Econom. 170, 83–97. http://dx.doi.org/10.1016/j.jpubeco.2019.01.008.

RKI, 2015. Gesundheit in Deutschland. Gesundheitsberichterstattung des Bundes. Gemeinsam getragen von RKI und Destatis. Technical Report, Robert Koch-Institut (Hrsg), Berlin.

Roth, Jonathan, Sant'Anna, Pedro H.C., 2023. When is parallel trends sensitive to functional form? Econometrica 91 (2), 737–747.

Sant'Anna, Pedro H.C., Zhao, Jun, 2020. Doubly robust difference-in-differences estimators. J. Econometrics 219 (1), 101–122.

Schwartz, Aaron L, Landon, Bruce E, Elshaug, Adam G, Chernew, Michael E, McWilliams, J Michael, 2014. Measuring low-value care in medicare. JAMA Internal Med. 174 (7), 1067–1076.

Simon, Michael, 2017. Das Gesundheitssystem in Deutschland. Hogrefe.

Van Winssen, K.P.M., Van Kleef, R.C., Van de Ven, WPMM, 2016. Potential determinants of deductible uptake in health insurance: How to increase uptake in The Netherlands? Eur. J. Health Econom. 17, 1059–1072.

Zweifel, Peter, 1987. Bonus systems in health insurance: a microeconomic analysis. Health Policy 7 (2), 273–288. http://dx.doi.org/10.1016/0168-8510(87)90037-6.

Zweifel, Peter, 1992. Bonus Options in Health Insurance. Springer Science & Business Media, Dordrecht, The Netherlands.