DISCUSSION PAPER SERIES

DP14552
(v. 2)

The Response to Dynamic Incentives in Insurance Contracts with a Deductible: Evidence from a Differences-in-Regression-Discontinuities Design

Tobias Klein, Martin Salm and Suraj Upadhyay

LABOUR ECONOMICS

CEPR

ISSN 0265-8003

The Response to Dynamic Incentives in Insurance Contracts with a Deductible: Evidence from a Differences-in-Regression-Discontinuities Design

Tobias Klein, Martin Salm and Suraj Upadhyay

Discussion Paper DP14552
First Published 02 April 2020
This Revision 29 September 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

The Response to Dynamic Incentives in Insurance Contracts with a Deductible: Evidence from a Differences-in-Regression-Discontinuities Design

Abstract

We develop a new approach to quantify how patients respond to dynamic incentives in health insurance contracts with a deductible. Our approach exploits two sources of variation in a differences-in-regression-discontinuities design: deductible contracts reset at the beginning of the year, and cost-sharing limits change over the years. Using rich claims-level data from a large Dutch health insurer we find that individuals are forward-looking. Changing dynamic incentives by increasing the deductible by 100 euros leads to a reduction in healthcare spending of around 3% on the first days of the year and 6% at the annual level. We find that the response to dynamic incentives is an important part of the overall effect of cost-sharing schemes on healthcare expenditures—much more so than what the previous literature has suggested.

JEL Classification: I13, H51

Keywords: Patient cost-sharing, Health Insurance, Dynamic incentives

Tobias Klein - t.j.klein@uvt.nl
Tilburg University and CEPR

Martin Salm - m.salm@uvt.nl
Tilburg University

Suraj Upadhyay - s.upadhyay@uvt.nl
Tilburg University

Acknowledgements
We would like to thank Sara Abrahamson and Joachim Winter as well as conference participants at the 2018 EuHEA conference in Maastricht, the 2019 Essen Health Conference, the 2019 iHEA conference in Basel, and the 2019 Meeting of the Health Economics Committee of the German Economic Association in Munich for helpful comments and suggestions. This paper is a follow up to work on the PACOMED project on patient cost sharing that was funded by the Netherlands National Institute for Public Health and the Environment (RIVM). There are no conflicts of interest.

The Response to Dynamic Incentives in Insurance Contracts with a Deductible: Evidence from a Differences-in-Regression-Discontinuities Design*

Tobias J. Klein, Martin Salm, and Suraj Upadhyay†

September 2021

Abstract

We develop a new approach to quantify how patients respond to dynamic incentives in health insurance contracts with a deductible. Our approach exploits two sources of variation in a differences-in-regression-discontinuities design: deductible contracts reset at the beginning of the year, and cost-sharing limits change over the years. Using rich claims-level data from a large Dutch health insurer we find that individuals are forward-looking. Changing dynamic incentives by increasing the deductible by €100 leads to a reduction in healthcare spending of around 3% on the first days of the year and 6% at the annual level. We find that the response to dynamic incentives is an important part of the overall effect of cost-sharing schemes on healthcare expenditures—much more so than what the previous literature has suggested.

Key words: Patient cost-sharing, health insurance, dynamic incentives.

JEL-classification: I13, H51.

*We would like to thank Sara Abrahamson and Joachim Winter as well as conference participants at the 2018 EuHEA conference in Maastricht, the 2019 Essen Health Conference, the 2019 iHEA conference in Basel, and the 2019 Meeting of the Health Economics Committee of the German Economic Association in Munich for helpful comments and suggestions. This paper is a follow up to work on the PACOMED project on patient cost sharing that was funded by the Netherlands National Institute for Public Health and the Environment (RIVM). There are no conflicts of interest.

†Klein: Tilburg University, Department of Econometrics and Operations Research; e-mail: T.J.Klein@uvt.nl; Salm: Tilburg University, Department of Econometrics and Operations Research; Upadhyay: Tilburg University, Department of Econometrics and Operations Research.

1 Introduction

Annual deductibles are a common feature in health insurance contracts in many countries, as well as in other types of insurance contracts. In the United States, 83% of employer-sponsored health insurance plans feature an annual deductible (Kaiser Family Foundation, 2020). In the Netherlands, annual deductibles are mandatory by law in health insurance contracts for adults. With deductibles, patients have to pay for a specific amount of healthcare out-of-pocket before insurance coverage begins, which gives rise to both static and dynamic incentives. Static incentives exist when patients have to make deductible payments for current healthcare use. Dynamic incentives exist when deductible payments for current healthcare utilization reduce deductible payments for healthcare utilization later in the year. Decisions about healthcare utilization will be affected by dynamic incentives if patients are aware of them, are forward-looking, and value future states sufficiently. Consider for instance the case of a patient with an expensive chronic disease, who will cross the limit for the annual deductible (almost) certainly. Such a patient will pay the full deductible amount anyway and therefore her current healthcare use should not (or should barely) respond to the deductible.

From a policy perspective, knowledge of whether and how individuals respond to dynamic incentives imposed by deductibles (or similar cost-sharing schemes) is crucial for the design of health insurance contracts: setting the amount of a deductible is an essential element in the design of many health insurance contracts, and understanding how people respond to insurance contracts with different deductible limits requires knowledge of how people respond to dynamic incentives. For example, forward-looking patients will respond less to a deductible than myopic (not forward-looking) patients. The reason is that forward-looking patients anticipate that higher out-of-pocket spending now might decrease future out-of-pocket spending while myopic patients do not take this into account. Related to this, it is important to understand whether different groups in the population respond differently to dynamic incentives because this will lead to differences in utilization even if healthcare needs are the same.

It is a challenge to quantify the reaction to dynamic incentives in this context because doing so requires a setting where dynamic incentives vary while all other factors—including static incentives—are kept constant. Previous studies have taken two alternative approaches: a reduced-form approach, where quasi-experimental sources of variation are used to test whether individuals respond to dynamic incentives, or structural modeling, where the response to dynamic incentives is quantified using a fully specified structural model. Studies that use reduced-form methods (e.g., Aron-Dine et al., 2015, and Guo and Zhang, 2019) deliberately abstain from interpreting the magnitude of the effect they estimate, as the estimation equation is not derived from a structural model. Studies that use structural models (e.g., Einav et al., 2015, and Dalton et al., 2020) rely on functional form and distributional assumptions. An additional challenge researchers face when pursuing a structural approach is the difficulty in accounting for unobserved heterogeneity in healthcare needs and their persistence over time.

The aim of this paper is to quantify the response to dynamic incentives in the context of health insurance with a deductible. For this, we propose an approach that (i) combines advantages of the reduced form and the structural approach, and (ii) can be pursued in other institutional contexts, provided that healthcare usage data are available at a high frequency and deductible limits exogenously vary across multiple years. Related to (i), we combine the advantages of the reduced form and the structural approach by estimating a micro-founded reduced-form equation exploiting quasi-experimental variation. As a consequence of deriving the equation from a full structural model, we can interpret the magnitude of the effects we estimate in terms of utility parameters and (to some extent) perform counterfactual experiments. At the same time, by abstaining from estimating a full structural model and focusing on estimating one key parameter of one structural equation, we can make very explicit which variation we exploit. Related to (ii), our approach is applicable in many contexts, because it uses two standard features of health insurance contracts with a deductible. The first feature is that deductibles reset at the beginning of the year. The second feature is that dynamic incentives change across years, because the deductible limit changes. Changes in deductible limits across years are often mandated by law (e.g. in the Netherlands) or originate in changes in employer policy for employment based health insurance in the United States (e.g. Brot-Goldberg et al., 2017). We show that the combination of these two features—deductibles resetting at the beginning of the year and changes in deductible amounts over time—gives rise to a differences-in-regression-discontinuities design.

In our analysis, we proceed in two steps. In the first step, we follow individuals who have crossed the deductible in year y as they experience a reset in the deductible at the beginning of year $y+1$. This means that they do not face cost-sharing incentives at the end of year y, but face them at the beginning of year $y+1$. The change in care consumption between the end of year y and the beginning of year $y+1$ is local and driven by the change of static *and* dynamic incentives. We repeat this approach for other year-pairs. Importantly, by estimating changes locally, around the turn of the year, we control for seasonality in healthcare needs and individual heterogeneity. Moreover, by design, the change in static incentives around the turn of the year is the same for all year-pairs. This means that *differences* in regression discontinuities are directly related to *differences* in dynamic incentives across years.

In the second step, we relate the sizes of the estimated discontinuities to a commonly used measure of dynamic incentives, the expected end-of-year price at the beginning of the second year of the respective year-pair. The average expected end-of-year price varies across years due to changes in the deductible amount. A lower expected end-of-year price makes it more attractive to consume more care earlier in the year. Forward-looking behavior would imply a negative relationship between the changes of care consumption we estimate around the turn of the year and the expected end-of-year price at the beginning of the second year of the respective year-pair. To test this prediction, we apply both parametric and non-parametric statistical tests. Moreover, we estimate the size of the dependence between the discontinuities and our measure of dynamic incentives, and we give these estimates a structural interpretation.

We implement this approach using administrative data from a Dutch health insurer for the years 2008 to 2015. The population of insured individuals for which we have data is broadly representative of the Dutch population. It is not limited to certain groups such as the elderly or employees of a specific firm. The Netherlands provides a favorable setting for our purpose. Health insurance coverage and the set of providers patients have access to is comparable across insurance companies. Furthermore, cost-sharing incentives are the same across insurance providers. Deductibles are mandatory for all health insurance contracts for adults, and the minimum deductible amount is set by the Dutch government each year. Deductible amounts are low, which means that dynamic incentives are important. At the same time, deductibles have increased substantially over our study period, from €150 in 2008 to €375 in 2015. This provides variation in dynamic incentives that we exploit with our differences-in-regression-discontinuities approach.

We find that individuals respond to the expected end-of-year price, in a way that is consistent with forward-looking behavior. An increase in the expected end-of-year price reduces daily expenditures at the beginning of the year. Increasing the deductible by €100 leads to a reduction of around 3.0% in daily healthcare expenditures at the beginning of the year when patients have not yet exceeded their deductible. We show that our results are robust to alternative specifications and discuss a wide range of alternative explanations for our findings. These are changes to static incentives, strategic timing of medical care use, endogenous selection into our sample based on forward-looking behavior, and the salience of deductible sizes. We present evidence that speaks against these alternative explanations.[1]

We also explore whether the reaction to dynamic incentives differs across subgroups in our sample, defined by age, gender, and income. We first control for differences in healthcare needs and find that almost every subgroup exhibits forward-looking behavior. The only exception is the group of individuals that are below age 45. For the remaining groups, we find the reactions are very similar. An increase in the size of the deductible by €100 leads to a percentage change in daily expenditures that ranges from -2.5% to -3.5%. Then, we explore whether the reaction depends on healthcare needs. For subgroups defined by different predicted sickness measures, such as risk scores in year y, we find that a €100 increase in the deductible size leads to a very similar percentage change in daily expenditures as for the baseline sample.

Finally, we show that our framework allows us to conduct interesting counterfactual experiments that incorporate prior knowledge about reactions to static incentives from the literature. We find that increasing the deductible by €100 leads to a reduction in per capita yearly expenditures by approximately €211, which is around 10.55% of total yearly expenditures that count towards the deductible. Of this total effect, approximately €120 or 6% of yearly expenditures

[1]Our results in Section 7.1.3 suggest that a simple version of salience theory, where patients are generally less inclined to consume care when deductible amounts are higher, cannot explain our results. At the same time, we cannot rule out that more refined versions of the salience theory at least partly explain our results. We discuss this further in the concluding Section 9.

can be attributed to dynamic incentives, and the remainder to static incentives. Thus, the reduction due to dynamic incentives is large compared to the reduction due to static incentives. Generalizing our findings to the Dutch population would imply that the reaction to dynamic incentives alone leads to a reduction in expenditures by approximately €2 billion when the deductible size is increased by €100.

Our study relates to the literature on patients' responses to cost-sharing incentives (see surveys by Cutler and Zeckhauser, 2000, McGuire, 2011, Einav and Finkelstein, 2018). Earlier theoretical contributions have examined patients' responses to dynamic incentives under a health insurance contract with a deductible. They conclude that, under some assumptions, forward-looking individuals should only respond to the expected end-of-year price and not to static incentives (Keeler et al., 1977, Ellis, 1986). This forms the basis for using the end-of-year price as a measure of dynamic incentives. We discuss this literature in detail in Online Appendix A, where we also provide the micro foundation for our analysis.

Several recent empirical studies examine whether and how patients respond to dynamic incentives in the context of patient cost-sharing. For Medicare Part D, some studies test the hypothesis of full myopia (i.e., no reaction to dynamic incentives) through the estimation of a discount factor; a discount factor of 0 would indicate full myopia. Einav et al. (2015) utilize the non-linearity in prices caused by the donut hole structure of Medicare Part D plans to estimate a weekly discount factor of around 0.96, or 0.12 at the annual level, rejecting the hypothesis of full myopia. Dalton et al. (2020), on the other hand, estimate a discount factor of 0—an indication of complete myopia. Similarly, Abaluck et al. (2018) find evidence for substantial myopia.

In the setting of employee insurance in the United States, Brot-Goldberg et al. (2017) find that for high deductible health insurance plans dynamic incentives play only a minor role in determining healthcare utilization. Similarly, Guo and Zhang (2019) study the spending patterns of individuals who have a large expenditure planned in the future (childbirth). Their results are consistent with individuals exhibiting myopic behavior. Aron-Dine et al. (2015) focus on employees that enroll into health plans in different months within the same year. Under the assumption that individuals who enrolled into these health plans in different months are comparable, they relate differences in healthcare utilization to the differences in the expected end-of-year price that are driven by differences in the time at which individuals enrolled during the year. Using a differences-in-differences approach, they reject the hypothesis of full myopia and conclude that dynamic incentives do matter for healthcare utilization.

We contribute in various ways to the literature. On the substantial side, we show that a broad population of individuals strongly react to dynamic incentives and that changing these incentives through increasing deductible limits has quantitatively important effects. Moreover, we show that the reaction to dynamic incentives is similar across groups in the population, with the exception of young individuals.

In addition, our study makes two methodological contributions. First, we show that a combination of a key feature of standard deductible contracts, namely that they reset at the turn of the year, and exogenous variation in the deductible amount across years give rise to a differences-in-regression-discontinuities design. We demonstrate how this can be used to estimate the effects of dynamic incentives on patient behavior.

Second, we explicitly derive our estimation equation as a reduced form from an economic model. This allows us to give a structural interpretation to the response to dynamic incentives at the beginning of the year that we estimate, and to perform counterfactual experiments. Moreover, we show that the expected end-of-year price is either a valid measure or a good proxy of dynamic incentives for a broad class of models that includes the model by Keeler et al. (1977) and more recent models with quasi-hyperbolic discounting, such as the model by Abaluck et al. (2018).

Our study continues as follows: we describe the institutional background in Section 2 and our empirical approach in Section 3. Then, in Section 4, we provide details on the data. We discuss our empirical implementation in Section 5. The results of our main analysis are presented in Section 6. In Section 7 we discuss potential alternative explanations of our findings, and we present robustness checks. In Section 8 we perform a counterfactual experiment and quantify the effect of an increase of the deductible on annual expenditures, and the contribution of dynamic incentives to this effect. Section 9 concludes.

2 Institutional background

In the Netherlands, health insurance is mandatory. Individuals have to buy insurance from one of several competing health insurers. Insurance is funded in about equal proportions by income-dependent employer contributions and premiums paid by the insured. In addition, there is a risk equalization system between insurance providers. Insurers cannot base premiums on individual health risk,[2] and they cannot deny coverage for the basic package.

The basic package includes a wide range of services, such as general practitioner (GP) services, specialist and hospital care, prescription drugs, mental health care, and medical devices such as hearing aids and prostheses. The contents of the basic package are determined by law, and they are adjusted annually.[3] On top of the basic package individuals can purchase additional coverage on the market for supplemental health insurance (e.g., dental care). In our study we focus on care included in the basic package.[4]

Since 2008, by law health insurance plans have to feature a deductible for all residents who

[2]Group discounts in premiums of up to 10% are allowed.

[3]A list of the changes can be found (in Dutch) on https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/aanvullende%20onderzoeksbeschrijvingen/pakketwijzigingen-zorgverzekeringswet.

[4]The deductible applies only to care included in the basic package. It is not allowed to offer insurance for deductible payments.

are at least 18 years old.[5] Knowledge about the esistence of a deductible is almost universal in the Netherlands.[6] For each calendar year, the minimum, mandatory deductible amount is set by the Dutch Government to a baseline amount. Individuals are allowed to opt for a higher deductible.[7] There have been substantial increases in the amount of the mandatory deductible over time: for example, the deductible was €150 in 2008; in 2015, it was €375.[8] The deductible resets annually, regardless of how much healthcare was consumed in the previous year. Some services are exempt from the deductible, such as consults by General Practitioners (GPs), maternity care, and medical equipment on rent (e.g., wheelchairs).

Treatments are largely billed in terms of diagnosis treatment combinations (DTCs). A DTC compensates for all care administered within an episode of treatment, including follow-up visits. Compensation for DTCs is determined through bargaining between insurers and providers. Providers send a bill to health insurers, who then determine how much patients have to pay out-of-pocket depending on their remaining deductible. Patients make deductible payments to the insurer, not to providers.

3 Empirical approach

3.1 Financial incentives and care consumption

The aim of this paper is to measure patients' responses to dynamic incentives. We do so in the context of a deductible contract for health insurance, where a patient pays for the first euros of care consumption herself and faces no out-of-pocket payments after exceeding the deductible limit. A key institutional feature of almost all deductible contracts is that they reset at the beginning of the calendar year. In this section, we explain how our differences-in-regression-discontinuities approach exploits this to identify the reaction to dynamic incentives and to disentangle it from the reaction to static incentives.

It is useful to think of deductible contracts in terms of prices. The current price of individual i in period t is defined as the price of the last unit of care in this period. The current price depends on whether or not an individual has exceeded the deductible limit at a given point in time, and it reflects static incentives in deductible contracts. If it is 1, then additional care has to be paid

[5] We exploit this feature to conduct a placebo analysis for children below age 18.

[6] Survey evidence suggests that knowledge about the existence of the deductible has spread quickly in the Dutch population since its introduction in the year 2008. While immediately after the introduction in the year 2008 only 42% of the population knew about the deductible, this share has increased to 95% in the year 2009 and 97% in the year 2010 (Ecorys, 2011).

[7] This voluntary deductible can be up to €500 above the mandatory deductible. Less than 4 percent of the individuals in our data choose a voluntary deductible, and we omit them from our sample.

[8] The mandatory deductible was €155 in 2009, €165 in 2010, €170 in 2011, €220 in 2012, €350 in 2013 and €360 in 2014. The increases in deductibles in the years since 2010 reflect the political preferences of governments under prime minister Rutte, head of the centre-right VVD party. The size of deductibles is a contentious political topic in the Netherlands, with more right-wing parties in favor of higher deductibles and left-wing parties in favor of lower (or no) deductibles. The mandatory deductible for the following year is announced with the budget on "prince's day" on the third Tuesday of September.

out-of-pocket. If it is 0, then additional care is free.

As noted already by Keeler et al. (1977), when deciding how much care to consume in t, individuals should take into account that out-of-pocket spending today can be seen as an investment that is associated with what they call a bonus: in expectation, out-of-pocket spending today will lower the price of care tomorrow. Based on a contribution by Ellis (1986), who characterizes optimal behavior for that model, a commonly used measure of dynamic incentives in this context is the expected end-of-year price.[9] For the standard deductible contracts we consider in this paper, at any point in time, this is equal to the probability that the patient will have to pay for the last unit of care in the year. This is also the measure we use. So, our aim is to measure how care consumption depends on the probability to pay out-of-pocket for the last unit of care in the year, controlling for medical needs and the current price.

Denote current prices by P^c_{it} and the expected end-of-year price by P^e_{it}. The superscript "c" stands for current and the superscript "e" stands for expected. Denote healthcare consumption of individual i in period t as c_{it} and baseline consumption as κ_{it}. Our reduced-form estimation equation makes the dependence of healthcare consumption on medical needs and both prices explicit by writing it as the sum of three parts:

$$c_{it} = \kappa_{it} - \gamma^c \cdot P^c_{it} - \gamma^e \cdot P^e_{it}. \tag{1}$$

So, κ_{it} is the consumption of care when care is free. Our aim is to estimate the dependence of care consumption on dynamic incentives, γ^e.

3.2 Micro foundation

The reduced form equation (1) can be derived from a dynamic structural model of healthcare consumption. The advantage to providing such a structural foundation is that it gives the parameters γ^c and γ^e in the reduced form equation (1) a structural interpretation. Moreover, it justifies basing counterfactual experiments on estimates of the parameters of the reduced-form equation, as we can think of the underlying structural relationship as being stable under policy variation.

We now briefly describe the model setup. Online Appendix A contains all derivations for a more general version of the model and a discussion of various technical points. It explains in detail why care consumption can be written as a function of the two prices. The appendix also discusses how the model relates to other models in the literature and in which sense the expected end-of-year price is a good measure of dynamic incentives.

In the model, patient i knows how much care she has consumed up to t. She learns in period t about her medical needs λ_{it} and forms expectations on the likelihood to hit the deductible limit by the end of the calendar year. Her flow utility is quasi-linear in money and quadratic in the

[9] See for instance Keeler and Rolph (1988), Aron-Dine et al. (2015), Brot-Goldberg et al. (2017), and Abaluck et al. (2018). See Online Appendix A and in particular Online Appendix A.5.1 for a detailed discussion.

difference between care consumption and medical needs. The utility function is specified such that if a patient has to pay for care consumption in the last period of the year (so that choice is static), then care consumption will be equal to the medical need λ_{it}. Conversely, if care is free, then patients consume $\lambda_{it} + \omega$. This means that ω is the additional care patients consume when it is free, which can be interpreted as a price effect and a measure of moral hazard.

Patients are quasi-hyperbolic discounters (O'Donoghue and Rabin, 1999) and dynamically optimize. β is a measure of present bias. It is between 0 and 1. If patients are fully aware of dynamic incentives, are forward-looking, and are not present-biased, then $\beta = 1$. Otherwise, $\beta < 1$.

We show in Online Appendix A that the parameters γ^c and γ^e in (1) are functions of ω and β:

$$\gamma^c = \omega \cdot (1 - \beta)$$
$$\gamma^e = \omega \cdot \beta.$$

This means that the more individuals are present biased (the smaller β) the more they react to static incentives (measured by the current price) and the less they react to dynamic incentives (measured by the expected end-of-year price). The sum of the two reactions is always equal to our overall measure of moral hazard, ω.

The aim of this paper is to estimate the response to dynamic incentives, $\gamma^e = \omega \cdot \beta$. This response is stronger if patients exert more moral hazard (ω is higher) and/or if they are less present-biased (β is closer to 1). Our aim is not to separately estimate β and ω. Instead, we show in Section 8 that we can perform counterfactual simulations based on an estimate of γ^e.

Next we show how we exploit the differences-in-regression discontinuities design to estimate γ^e.

3.3 Differences-in-regression-discontinuities design

A challenge for estimating the effect of prices on care consumption using individual-level data is that prices are endogenous: higher care consumption earlier in the year is associated with lower current prices and lower expected end-of-year prices, and at the same time likely positively correlated with medical needs later in the year. This means that prices will be negatively correlated with medical needs so that a regression of medical care on prices yields negatively biased coefficient estimates.

The approach we pursue in this paper is to identify γ^e by exploiting a differences-in-regression-discontinuities design that allows us to use variation in the average P^e_{it} at the beginning of the year. This variation is across years and driven by changes in the deductible amount. Our approach allows us to control for static incentives, individual heterogeneity, and time effects.

Denote averages over individuals in our samples in period t of year y by a bar indexed by t

and y, where periods t range from $t = 1$ for the first period up to $t = T$ for the last period in a given year y. Thus, $\bar{c}_{t,y}$ is the average care consumption in period t of year y, $\bar{P}^c_{t,y}$ is the average current price in period t of year y, and $\bar{P}^e_{t,y}$ is the average expected end-of-year price in period t of year y. Note that care consumption in (1) is a linear function of P^e_{it}. This implies that average care consumption is a function of the average expected end-of-year price.

We work with samples for year-pairs. The samples consist of observations for the last periods of the first year of the year-pair and the first periods of the second year. We restrict the samples to individuals who have reached the deductible limit after January and before September of the first year.[10] Thus, all individuals in our samples face a current price of 0 by the end of the first year of the year-pair, and therefore their expected end-of-year price is 0. Deductibles reset each year and therefore all individuals face a current price of 1 right at the beginning of the second year in the year-pair.[11] This is the same for all years. Importantly, the average expected end-of-year-price at the beginning of the second year in the year-pair is different across years due to changes in deductible amounts.

Figure 1 illustrates our approach with an example. The left side shows average spending around the turn of the year for individuals who have crossed the deductible in year y. The solid line denotes average spending before the turn of the year, when $\bar{P}^c_{T,y} = 0$ and $\bar{P}^e_{T,y} = 0$. When individuals enter the new year, they face a current price of $\bar{P}^c_{1,y+1} = 1$ and a positive average expected end-of-year price $\bar{P}^e_{1,y+1}$, and they respond to the increase in prices by reducing their average spending. The right side shows care consumption around the turn of the year from $y+1$ to $y+2$ for a comparable sample of individuals who have exceeded the cost sharing limit in $y+1$. For this group we have $\bar{P}^c_{T,y+1} = 0$ and $\bar{P}^e_{T,y+1} = 0$. In the figure, consumption just before the end of $y+1$ (in the right year-pair) is the same as just before the end of y (in the left year-pair). At the beginning of the year, the current price is again $\bar{P}^c_{1,y+2} = 1$ in $y+2$. However, the expected end-of-year price is higher than in $y+1$, $\bar{P}^e_{1,y+2} > \bar{P}^e_{1,y+1}$, because the deductible is higher (and hence it is less likely that comparable individuals exceed the cost sharing limit). Therefore, when they enter the new year, they face a higher $\bar{P}^e_{1,y+2}$ and reduce their spending by a larger amount than in $y+1$.

In the following, we formalize the intuition illustrated in Figure 1 and describe how we can estimate γ^e from differences in regression discontinuities. Formally, it follows from (1) that the discontinuity in care consumption around the turn of the year from year y to $y+1$ is

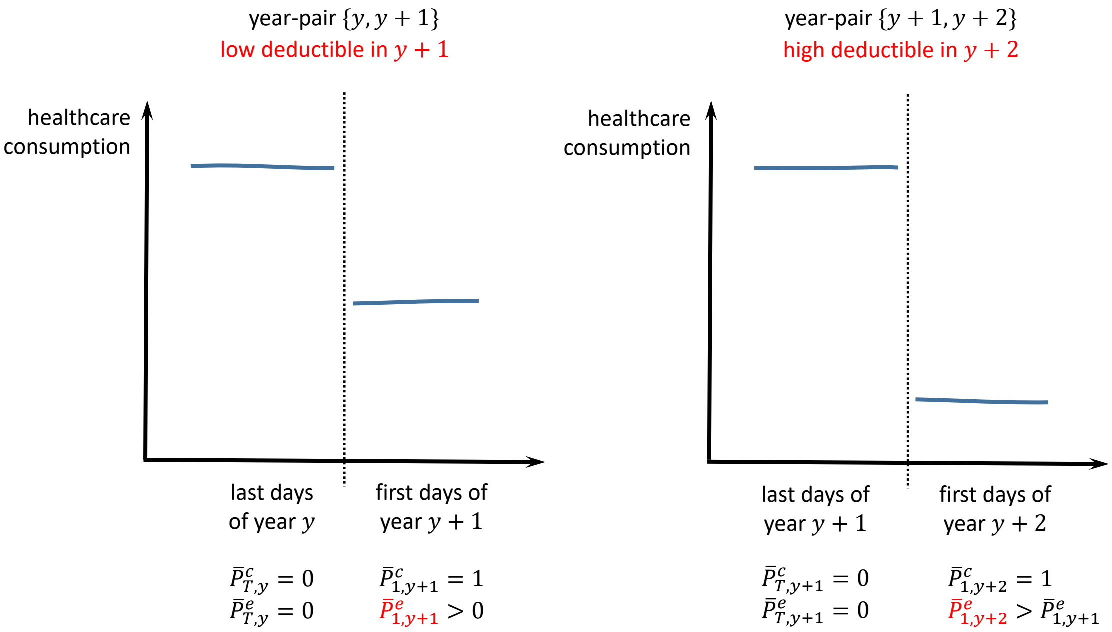
$$\bar{c}_{1,y+1} - \bar{c}_{T,y} = \bar{\kappa}_{1,y+1} - \bar{\kappa}_{T,y} - \gamma^c \cdot \bar{P}^c_{1,y+1} - \gamma^e \cdot \bar{P}^e_{1,y+1},$$

since our samples include only individuals for whom $\bar{P}^c_{T,y} = 0$ and $\bar{P}^e_{T,y} = 0$. For the discontinuity from year $y+1$ to $y+2$ we obtain a similar expression. Combining those, the difference in the

[10] We select the samples in a way so that the level of medical needs is comparable across year-pairs. Details are provided in Section 4 below.

[11] It is possible that some individuals already exceed the deductible limit in the first period of the year. We discuss this in Section 7.1.

Figure 1: Differences-in-regression-discontinuities design



Notes: The figure illustrates the intuition of our empirical approach for hypothetical year-pairs $\{y, y+1\}$ and $\{y+1, y+2\}$. The vertical lines depict the respective turn of the year. For both year-pairs the solid line depicts average care consumption at the end of the year (to the left of the vertical line) and average care consumption at the beginning of the new year (to the right of the vertical line). Individuals in the figure on the right reduce their average spending by a larger amount than individuals in the figure on the left because they face a higher expected end-of-year price at the beginning of the new year. The notation in the bottom is explained in the main text.

discontinuities is given by

$$(\bar{c}_{1,y+2} - \bar{c}_{T,y+1}) - (\bar{c}_{1,y+1} - \bar{c}_{T,y}) = (\bar{\kappa}_{1,y+2} - \bar{\kappa}_{T,y+1}) - (\bar{\kappa}_{1,y+1} - \bar{\kappa}_{T,y}) \\ - \gamma^c \cdot (\bar{P}^c_{1,y+2} - \bar{P}^c_{1,y+1}) - \gamma^e \cdot (\bar{P}^e_{1,y+2} - \bar{P}^e_{1,y+1}).$$

Since $\bar{P}^c_{1,y+2} = \bar{P}^c_{1,y+1} = 1$ at the beginning of the year,[12] we obtain

$$(\bar{c}_{1,y+2} - \bar{c}_{T,y+1}) - (\bar{c}_{1,y+1} - \bar{c}_{T,y}) = (\bar{\kappa}_{1,y+2} - \bar{\kappa}_{T,y+1}) - (\bar{\kappa}_{1,y+1} - \bar{\kappa}_{T,y}) - \gamma^e \cdot (\bar{P}^e_{1,y+2} - \bar{P}^e_{1,y+1}).$$

This equation differences out different levels in baseline care consumption across year-pairs. Such differences could arise, for example, because the flu was particularly severe in some winters. If the flu season was particularly severe around the turn of the year from y to $y+1$, then this would affect both $\bar{\kappa}_{T,y}$ and $\bar{\kappa}_{1,y+1}$. By taking differences between regression discontinuities we control for such differences in seasonal effects across year-pairs.

[12] Recall that it is possible that some individuals already exceed the deductible limit in the first period of the year (footnote 11). For our approach to be valid we only need $\bar{P}^c_{1,y+2} = \bar{P}^c_{1,y+1}$. In Section 7.1 we show supportive evidence for this.

Our main identifying assumption is that

$$(\bar{\kappa}_{1,y+2} - \bar{\kappa}_{T,y+1}) = (\bar{\kappa}_{1,y+1} - \bar{\kappa}_{T,y}) \tag{2}$$

This assumes that the difference in mean baseline consumption across the turn of the year is the same for both year-pairs. In Section 7, we discuss the plausibility of this assumption, and we conduct a number of robustness checks. Under this assumption, our parameter of interest is given by the ratio between the difference in regression discontinuities and the difference in expected end-of-year prices from the perspective of the beginning of the year,

$$\gamma^e = \frac{(\bar{c}_{1,y+2} - \bar{c}_{T,y+1}) - (\bar{c}_{1,y+1} - \bar{c}_{T,y})}{(\bar{P}^e_{1,y+2} - \bar{P}^e_{1,y+1})}.$$

The derivation so far was for two year-pairs. For multiple year-pairs, as we have in our study, we can obtain the discontinuity for each year-pair and then plot it against the average expected end-of-year price in the respective second year of each year-pair. We can use this as the basis for testing whether the relationship is monotonic and to estimate γ^e using a linear regression of the discontinuity on the expected end-of-year price. See Section 5.2.

4 Data

We use claims data from a large Dutch health insurer for the years 2008 to 2015 (see Hayen et al., 2015 for details).[13] We restrict our analysis to types of care that are part of the basic package and count towards the deductible.[14] Our data include the amount paid for a claim, the date of claim initiation, the type of claim, and demographic information, such as age and gender of the enrollees.

We construct separate samples for each year-pair. Each sample consists of individuals who cross the deductible at any point between February and August (inclusive) in year y and we follow these individuals around the turn of the year, from y to $y+1$.[15] Conditioning on individuals who cross the deductible in a given year leads to sicker samples in years with a larger deductible, as individuals have to spend more to be included in the sample of crossers for these

[13] Our data were obtained under a pilot project in which a new payment model for GPs was evaluated, The pilot project started in July 2014, and the data cover multiple years before the start of the pilot project. While the pilot project was not related to patient cost-sharing we cannot exclude the possibility that the results for the last year-pair {2014,2015} could be influenced by the treatment in the pilot project. In order to take this into account, we conduct a sensitivity analysis in which we restrict our analysis to year-pairs before the start of the pilot project. See Online Appendix D.

[14] This means that in our main analysis we do not look at consults at a GP or maternity care, for example.

[15] We exclude individuals that cross the deductible in January of year y since they have a very different pattern of healthcare expenditures when compared to the rest of the sample. We exclude individuals who cross the deductible after August of year y because healthcare expenditures can be autocorrelated over time, and we aim to limit the influence of healthcare expenditures associated with crossing the deductible in the previous year on expenditures across the turn of the year.

years: for example, in 2008, an individual would have to spend only €150 to cross the deductible, while in 2014 she would have to spend €360. To address such concerns, we utilize a percentile-matching strategy based on Brot-Goldberg et al. (2017). Specifically, across all years we include only the top 38% of cumulative spenders by the end of August in our sample. 38% is the share of individuals who exceed the deductible limit between February and the end of August in the year 2014, the year with the lowest share of such individuals in our data. In 2008, for example, even though 47% of the sample crossed the deductible between February and the end of August, we only include the top 38% of total spenders in our sample.[16] The total number of individuals in our final sample, after applying percentile-matching, and the total number of individuals without matching in each year are reported in the last two rows of Table 1.

Our goal is to compare healthcare utilization at the end of year y with healthcare utilization at the beginning of year $y+1$. For this, we specify the date of treatment as the date the claim was initiated and aggregate our claims data to the daily level.

Non-emergency care is limited in availability on weekends and during the Christmas break. Therefore, we only use week days for our main analysis.[17] Moreover, for each year-pair we omit a set of days that include the Christmas break. For this, we specify a last regular day before the start of the Christmas break (we denote this day by $t = T$ in the first year of the year-pair) and a first regular day (we denote this day by $t = 1$ in the second year of the year-pair) after the end of the Christmas break. The resulting empirical setup is referred to as a donut regression discontinuity (RD) design (Barreca et al., 2011).[18] Gerfin et al. (2015) have used it earlier to estimate the size of the discontinuity in healthcare expenditures around the turn of the year for one year-pair.[19]

The distribution of healthcare expenditures is characterized by a heavy right tail. Although our sample size is not small, outliers can still have a large influence on estimated coefficients. To ensure that our results are not influenced by high healthcare expenditures far above the deductible amount, we pseudo-censor our expenditure variable. For this, we code any daily expenditure above €500 as €500. This cutoff is higher than the highest deductible in our study period. In Online Appendix D we show that our results are robust to changes in this cutoff amount and to an alternative transformation of the expenditure variable (logs instead of levels), adding a constant of 1 to account for the zeros present in the data. The distribution of

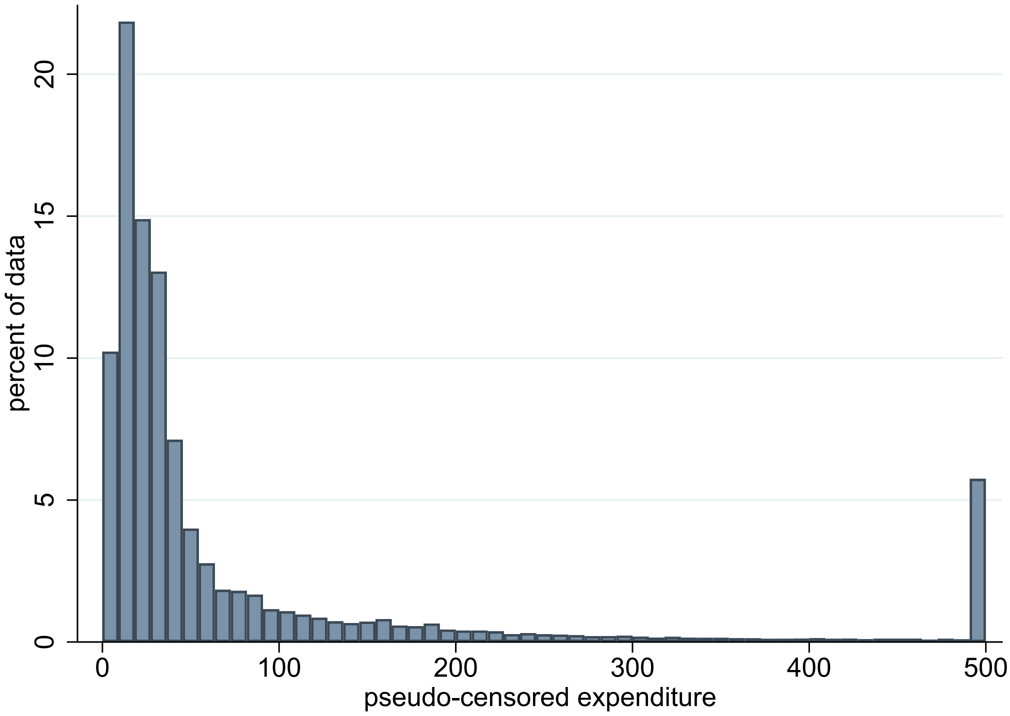
[16] The assumptions underlying this approach are the same as in Brot-Goldberg et al. (2017). We assume that there is (i) a monotone one-to-one mapping between healthcare needs and spending, and that (ii) the distribution of healthcare needs in the population is constant across years. Under these assumptions, the top 38% spenders in each year are comparable across years in terms of needs. In Section 7.1.4 we discuss possible violations of this assumption, and we present a robustness check.

[17] In Section 7.3 and Table A.4 in the Online Appendix we discuss and present alternative specifications that also include weekends.

[18] The dates we use in our preferred specification are detailed in Online Appendix C. We use a donut hole of 20 days across all year-pairs, except for the year-pair $\{2011, 2012\}$, where the donut hole is 27 days long. Section 7.3 and Table A.4 in the Online Appendix show the robustness of our primary results with respect to changes to the days—and the distance between the days—for which utilization changes are compared.

[19] They also document in detail, in their Figure 2, that care consumption is lower on weekends and during the Christmas break.

Figure 2: Distribution of daily expenditures (conditional on any spending)



Notes: The figure plots the histogram of pseudo-censored daily expenditures. We condition on any spending and pool across all years in our sample. The vertical axis shows the proportion of our data with expenditures falling within each bin. The mass point at €500 arises because we code expenditures larger than 500 as 500.

daily expenditures (conditional on any spending) is depicted in Figure 2. €500 is at the 95th percentile of this distribution.

Table 1 shows summary statistics, separately for all 7 year-pairs, $\{y, y+1\}$, in our data. The first three rows report the average age, proportion female, and average income at the 6-digit postal code level in each year y.[20] The next two rows show the deductible in year $y+1$ and the in-sample expected end-of-year price $\bar{P}^e_{1,y+1}$ at the beginning $(t=1)$ of year $y+1$.[21] Generally speaking, there is a positive relationship between the two, but this is not always the case: for example, between 2008 and 2009 the deductible increased by €10, while $\bar{P}^e_{1,y+1}$ decreased.

Table 1 also shows different measures of average healthcare utilization on regular days (i.e., excluding days in the donut hole and weekends), for the last four months in year y and the first month of year $y+1$. We look at mean daily expenditures, mean pseudo-censored daily expenditures, and the probability of having any daily expenditure. We see that utilization, across all measures, increases across years both at the end of year y and at the start of year $y+1$. We

[20]In comparison, average age in the general adult population is 48.75 years. The proportion female is 0.52. Average income is €2188. See the second column of Table 1 in Hayen et al. (2021).

[21]We compute the expected end-of-year price as $1-\Pr(cross_{y+1})$, where $\Pr(cross_{y+1})$ is the proportion of individuals in the sample who cross the deductible by the end of year $y+1$.

Table 1: Summary statistics

	year-pair $\{y, y+1\}$						
	2008, 2009	2009, 2010	2010, 2011	2011, 2012	2012, 2013	2013, 2014	2014, 2015
demographics in y							
average age	54.57	54.49	54.68	55.35	55.95	57.05	57.93
percentage female	.575	.575	.572	.565	.568	.566	.567
average income[a]	2123	2121	2141	2121	2123	2114	2114
dynamic incentives at the beginning of $y+1$							
deductible in $y+1$	155	165	170	220	350	360	375
expected end-of-year price $(\bar{P}^e_{1,y+1})^b$	.1452	.1443	.1534	.1691	.2170	.2124	.2200
average daily expenditure[c]							
September to December year y	11.74	13.14	12.72	13.94	13.03	14.60	14.09
January of year $y+1$	12.92	10.56	10.50	10.93	11.81	12.32	11.32
average daily expenditure (PC)[c,d]							
September to December year y	4.50	4.83	4.98	5.23	5.48	5.96	6.01
January of year $y+1$	4.41	4.76	4.55	4.78	5.11	5.38	5.31
prob. of any daily expenditure[c]							
September to December year y	.057	.062	.064	.069	.070	.076	.078
January of year $y+1$	.055	.063	.061	.064	.066	.074	.074
no. of individuals without matching[e]	29289	32953	33985	36329	33845	31282	29626
no. of individuals in estimation sample$(N)^f$	22798	25090	27473	26706	28045	30864	29626

Notes: This table shows summary statistics for our baseline estimation sample.
[a] Average income is at the 6-digit postal code level in 2008.
[b] $\bar{P}^e_{1,y+1}$ is computed as 1 minus the proportion of individuals in the sample who cross the deductible by the end of year $y+1$.
[c] We compute average utilization for regular days only, i.e., we do not include utilization around holidays (see Table A.1 in the Online Appendix) and weekends when computing this average.
[d] PC refers to "pseudo-censored". The threshold used to pseudo-censor was 500.
[e] This is the number of individuals who exceed the deductible between February and August in year y and are observed in year $y+1$.
[f] This is the number of individuals in our estimation sample. They exceed the deductible between February and August in year y, are observed in year $y+1$, and are the top 38% of cumulative spenders by the end of August (percentile matching). See Section 4 for details.

also see that spending at the start of year $y+1$ is generally lower than at the end of year y, and we see that the difference between spending at the end of year y and spending at the beginning of year $y+1$ is generally larger for year-pairs with a larger $\bar{P}^e_{1,y+1}$. These patterns are in line with the prediction of our model illustrated in Figure 1 that in year-pairs with a larger $\bar{P}^e_{1,y+1}$ healthcare consumption decreases more around the turn of the year .

5 Empirical implementation

5.1 Estimation of discontinuity sizes

For each year-pair, we estimate the change in care consumption around the turn of the year using separate local linear regressions before and after the turn of the year. Thereby, we allow for separate smooth time trends in care consumption before and after the change of the year. These could be related to the Christmas holiday season or the flu. They could also be driven by the gradual increase in spending in the first weeks of the new year that originates in more and more patients exceeding the deductible limit.[22]

To estimate the discontinuity sizes, we specify separate linear functions in time (one for year y and one for year $y+1$ of each year pair) and then weigh observations using a kernel such that more weight is put on observations that are closer to the change of the year.[23]

The specification we use for the functions are

$$\begin{aligned} c_{it} &= \alpha_0 + \gamma_0 (t - (T+1)) + \varepsilon_{it} \text{ for year } y \text{ and } t \leq T \\ c_{it} &= \alpha_1 + \gamma_1 (t-1) + \varepsilon_{it} \text{ for year } y+1 \text{ and } t \geq 1. \end{aligned}$$

Day $t = T+1$ in year y is the same as day $t = 1$ in year $y+1$. Therefore, $\alpha_1 - \alpha_0$ is care consumption on the first day of year $y+1$ minus an out-of-sample prediction of care consumption on the day following the last day of year y, using only data for year y. This is the quantity we wish to estimate.

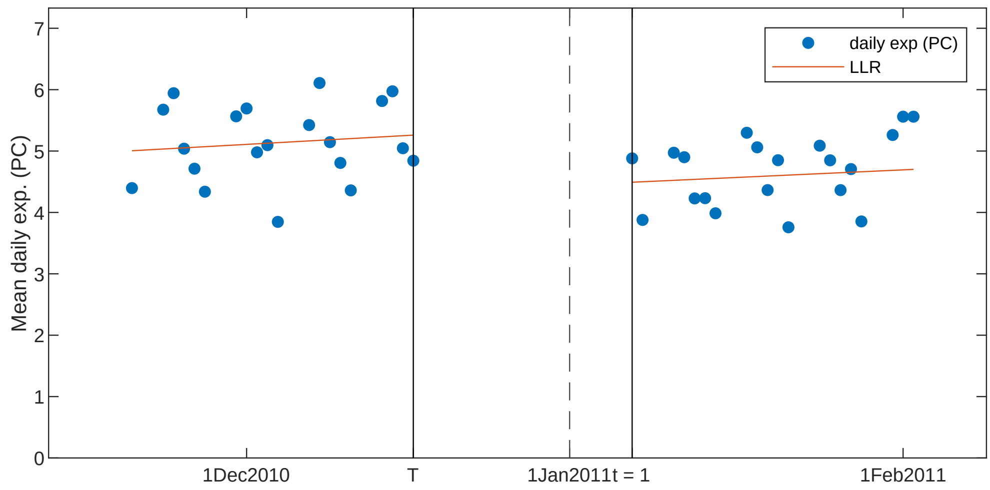
For the weighting, we use a triangular kernel.[24] This kernel puts positive weight on care consumption for the last h days of year y and the first h days of year $y+1$. h is called the bandwidth. For a triangular kernel, the weights start at 0 and increase linearly from 0 to 1 for the last $h+1$ days of year y; they decrease linearly from 1 to 0 for the first $h+1$ days for year

[22] We also performed the analysis with simpler specification in which we do not allow for time trends. Results are discussed in Section 7.3 and shown in Table A.4 and Figure A.21 in the Online Appendix.

[23] The general recommendation is to use at least a linear polynomial when one estimates a value of a function at the boundary, as we do. See Pagan and Ullah (1999) for details. We do not use a higher order polynomial, because Gelman and Imbens (2019) have recently shown that this may lead to problems in the context of estimating discontinuity sizes.

[24] The choice of kernel function does typically not have a big effect on the estimates. Asymptotically, it does not play a role, as long as one uses a kernel that satisfies the assumptions needed for the asymptotic analysis. This is the case for the triangular kernel. See for instance Pagan and Ullah (1999) for details.

Figure 3: Healthcare consumption around turn of year from 2010 to 2011



Notes: The figure plots mean daily pseudo-censored healthcare spending for the last 20 weekdays of 2010 and for the first 20 weekdays of 2011. T denotes the last day of the year 2010 that we use and $t = 1$ denotes the first day of the year 2011. These are the days we use for our analysis. They are Thursday 16 December, 2010, and Thursday 6 January, 2011 (see Table A.1 in the Online Appendix). The solid lines denote local linear regression estimates (LLR). See Section 5.1 for details on the estimation procedure. See Figure A.2 and Figure A.3 in the Online Appendix for the other year-pairs.

$y+1$. For the selection of the bandwidth and inference we follow Calonico et al. (2014).[25]

Figure 3 illustrates our approach for the year-pair $\{2010, 2011\}$. The last day that we use for our analysis, $t = T$ in 2010, is Thursday 16 December. The first day, $t = 1$ in 2011, is Thursday 6 January. The days in-between lie in the donut hole and are omitted from our analysis because non-emergency care is limited in availability on weekends and during the Christmas break.[26]

As described above, we follow the same individuals over time and estimate the discontinuity as the predicted value on 6 January 2011 (an in-sample prediction using the data from January 6 onward) minus the predicted value for 17 December 2010 (a one-day-ahead prediction using the data up to 16 December 2010).

There are two interesting aspects of our estimation procedure that we would like to highlight. First, since we follow a balanced panel of individuals, covariates are, by construction, balanced across both sides of the threshold we compare. Second, since we subtract predicted spending at the end of year y from predicted spending at the beginning of year $y+1$, we remove any effect of characteristics and influences that are invariant across these two dates, such as, for example, the severity of the flu season.

[25] We also performed our empirical analysis using bandwidths of 20 days and found almost no difference.

[26] In Section 7.3 and in Table A.4 in the Online Appendix we show that our results are robust to different specifications of the donut hole.

5.2 Relating discontinuity sizes to dynamic incentives

We repeat the above for all year-pairs, $\{y, y+1\}$. Our main focus is on estimating the relationship between the changes in expenditure around the turn of the year and $\bar{P}^e_{1,y+1}$. If patients are forward-looking, then the change should be bigger (in absolute terms) in year-pairs with higher $\bar{P}^e_{1,y+1}$.

We estimate our parameter of interest, γ^e, by regressing our estimates of the sizes of the discontinuities in care consumption around the turn of the year on $\bar{P}^e_{1,y+1}$. When doing so, it is important to take into account that not all discontinuity sizes are estimated equally precisely, as reflected in differences in the standard errors in Table 2.

The error term in this model consists by definition of two components. The first component comes from the fact that the linearity between the size of the discontinuity and $\bar{P}^e_{1,y+1}$ is an approximation. The second part comes from the estimation error in the first stage in which we estimate the size of the discontinuity. We assume that the variance of the first part is constant across observations and that the two parts are not correlated with one another. Under these assumptions, we can use a standard feasible generalized least squares procedure to obtain efficient estimates and correct standard errors (Hanushek, 1973). See Online Appendix B.1 for additional details.

This regression imposes a linear relationship between the estimated discontinuity size and $\bar{P}^e_{1,y+1}$. To test for monotonicity without imposing parametric restrictions, we follow Patton and Timmermann (2010) and use three non-parametric tests for monotonicity from the finance literature. All three tests rank the estimates of the discontinuities by their corresponding $\bar{P}^e_{1,y+1}$. We test whether the ordering of the $\bar{P}^e_{1,y+1}$ is the same as the ordering of the discontinuities. The MR test is significant if the size of the discontiunities monotonically increases with rank. The Up test is significant if the size of the discontinuities increases with rank for at least some ranks, and the Down test is significant if the size of the discontinuities decreases with rank for at least some ranks. We obtain p-values using a bootstrap procedure. See Online Appendix B.2 for details. The tests are constructed so that results aligning with forward-looking behavior produce p-values smaller than a chosen level of significance for the MR test and the Up test. The p-value for the Down test should be higher than the chosen level of significance. Patton and Timmermann (2010) show that the MR test may have low power, which can result in high p-values. We also find this for some of our specifications.

6 Results

6.1 Baseline results

In this section, we present our baseline results for two measures of healthcare utilization: mean (pseudo-censored) expenditures and the probability of any claim (extensive margin).

Table 2: Discontinuity sizes

year-pair $\{y, y+1\}$	change in daily expenditure (PC) around turn of the year	change in probability of any daily expenditure around turn of the year
2008,2009	-0.737	-0.010
	(0.1368)	(0.0010)
2009,2010	-0.473	-0.005
	(0.1252)	(0.0011)
2010,2011	-0.827	-0.008
	(0.1270)	(0.0009)
2011,2012	-0.965	-0.011
	(0.1255)	(0.0010)
2012,2013	-1.338	-0.016
	(0.1661)	(0.0014)
2013,2014	-1.296	-0.013
	(0.1345)	(0.0013)
2014,2015	-1.599	-0.013
	(0.1651)	(0.0011)

Notes: This table shows estimated changes in expenditures and the probability of any daily expenditure around the turn of the year, by year-pair. Expenditures larger than 500 were coded as 500 (hence the abbreviation PC). Changes are estimated using a donut hole regression discontinuity design (see Section 5.1). Robust standard errors that are clustered at the individual and year-pair level are shown in parentheses.

Table 3: Dependence of discontinuity sizes on dynamic incentives

	daily expenditure	any daily expenditure
effect of dynamic incentives (γ^e)	-10.762	-0.094
	(1.4883)	(0.0230)
p-value nonparametric MR test	0.000	0.927
p-value nonparametric Up test	0.004	0.000
p-value nonparametric Down test	1.000	0.173

Notes: This table presents our estimate of the effect of dynamic incentives and the results for three nonparametric tests for the monotonicity of the relationship between the size of the discontinuity in care consumption and $\bar{P}^e_{1,y+1}$. See Section 5.2for details.

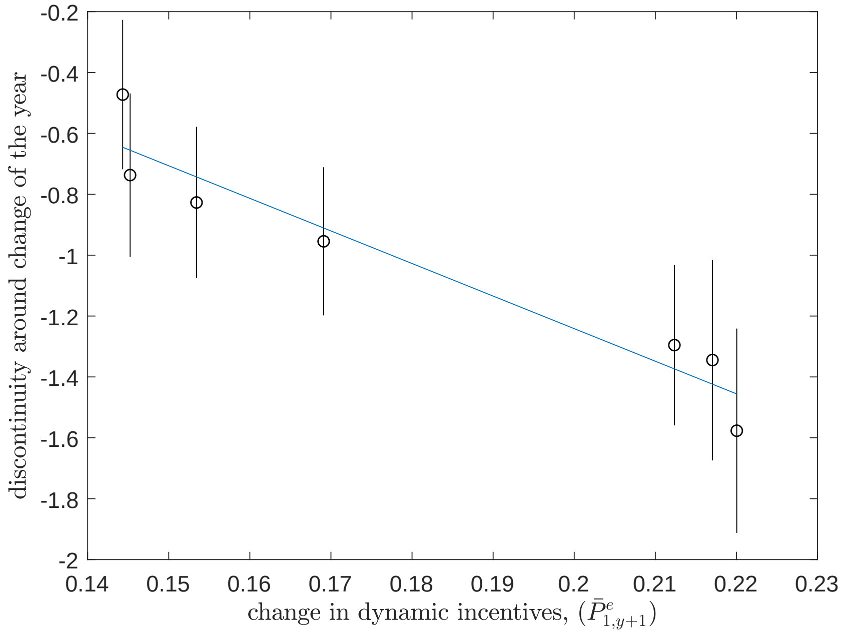
Table 2 reports the estimated changes in healthcare utilization at the turn of the year, for both measures. For all year-pairs, healthcare utilization decreases at the turn of the year, and these decreases are statistically significant at any conventional level. Decreases in healthcare utilization tend to be stronger for later years. For example, for the year-pair {2008,2009} the decrease in pseudo-censored daily expenditures is €0.74, and for the year-pair {2014,2015} the decrease is €1.60. Correspondingly, the probability of any daily expenditure decreases by 1.0 percentage point for the year-pair {2008,2009} and by 1.3 percentage points for the year-pair {2014,2015}.

Figure 4 plots these estimated changes in healthcare consumption at the turn of the year against $\bar{P}^e_{1,y+1}$. As we have seen in Table 1, $\bar{P}^e_{1,y+1}$ tends to be higher for later years in our sample when the deductible is higher. For example, $\bar{P}^e_{1,2009}$ is 0.15, while $\bar{P}^e_{1,2015}$ is 0.22. Figure 4 shows a decreasing relationship, providing evidence for forward-looking behavior. Table 3 reports our estimate of the dependence of the discontinuity size on dynamic incentives. Based on a weighted linear regression we find that the slope is equal to -10.76 for daily expenditures, and it is statistically significant at any conventional level. This coefficient is our estimate of γ^e. Our results thus suggest that increasing $\bar{P}^e_{1,y+1}$ by 10 percentage points leads to a reduction in mean (pseudo-censored) expenditures of around €1.08 and a reduction in the probability of any claim of around 0.9 percentage points.

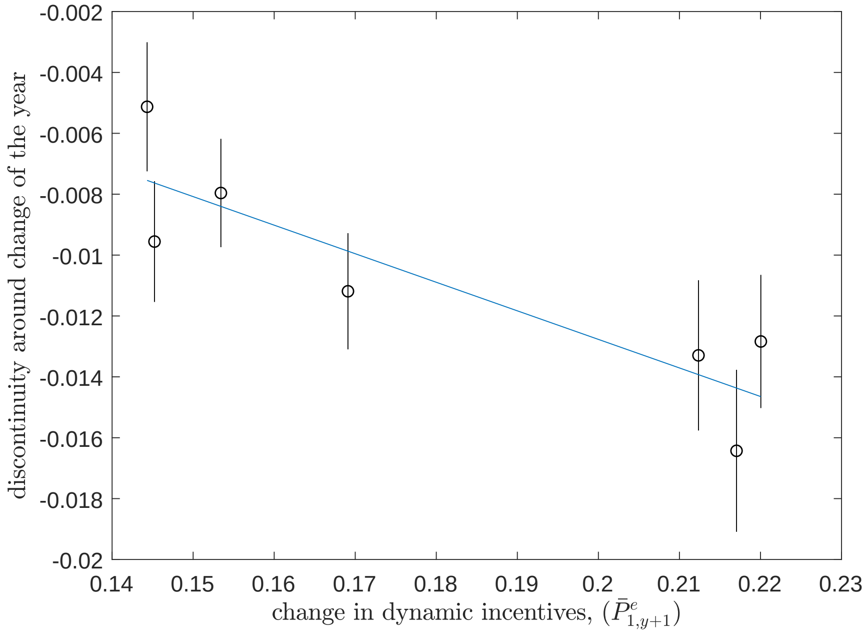
The bottom three rows of Table 3 show the results from the non-parametric tests for monotonicity in Section 5.2. Recall that results aligning with forward-looking behavior produce p-values smaller than a chosen level of significance for the MR test and the Up test. At the same time, the p-value for the Down test should be higher than the chosen level of significance. For pseudo-censored expenditures, we see that the MR test and the Up test have p-values below conventional levels and the Down test above it, indicating a monotonically decreasing relationship between mean expenditures and $\bar{P}^e_{1,y+1}$. For the extensive margin, the p-value of the MR test is above conventional levels of significance. However, the p-values from the Up and Down test suggest that this test has low power (as discussed in Section 5.2) and that there is a

Figure 4: Dependence of discontinuity sizes on dynamic incentives

(a) Daily spending (pseudo-censored)



(b) Prob. any daily expenditure



Notes: These figures plot the relationship between the estimated changes in healthcare consumption around the turn of the year and $\bar{P}^e_{1,y+1}$. See Tables 1 and 2 for exact numbers. The vertical lines denote 95% confidence intervals. The solid line is the OLS regression line.

decreasing relationship.

The variation in $\bar{P}^e_{1,y+1}$ comes from changes in deductibles. Therefore, we can relate changes in expenditures to changes in the deductible amount. We do so in the first row of Table 4. The third-to-last column shows that a €100 increase in the deductible leads to a 3.4 percentage point increase in $\bar{P}^e_{1,y+1}$.[27] This in turn implies a reduction of around €0.36 in mean (pseudo-censored) expenditures and a 0.32 percentage point reduction in the probability of making any claim (calculated by multiplying the effects in the first two columns by the estimated 3.4 percentage point changes in $\bar{P}^e_{1,y+1}$). In percentage terms, these reductions amount to around 3.0% of mean expenditures and 4.4% of the probability of any claim on the first day of the year.[28]

These calculations show that the relative reduction in the probability of any claim (4.4%) is larger than the relative reduction in mean expenditures (3%). Thus, the average size of claims increases in response to dynamic incentives. Put differently, at the extensive margin the effect is negative, while it is positive at the intensive margin.

6.2 Heterogeneity

Next, we examine whether and how responses to $\bar{P}^e_{1,y+1}$ differ for different groups in the population, defined by gender, neighborhood income, age, and predicted sickness. Previous studies such as Manning et al. (1987), Farbmacher et al. (2017), and (Hayen et al., 2021) have examined heterogeneous responses to cost-sharing incentives for different groups in the population. However, there is still limited evidence on heterogeneous responses specifically to dynamic incentives.

Responses to dynamic incentives can differ between groups as a result of two mechanisms that may be at play at the same time. The first mechanism is that the size of the reaction is related to healthcare needs, and that different groups in the population have different healthcare needs. This is plausible, as for instance the possibilities to exert moral hazard may be greater for older patients who are less healthy. The second mechanism is that different groups in the population respond differently to dynamic incentives, even if the needs are the same. This could be explained by differences in the willingness or ability to acquire and process information, and differences in utility parameters. For example, patients might respond less strongly (or not at all) to dynamic incentives if they are not aware of them, are myopic, or strongly discount the value of future states (a low β in the terms of our model). They might also respond less strongly to dynamic incentives if they exert less moral hazard (a low ω in the terms of our model).

In Table 4, we focus on the second mechanism. We examine how different groups in the population respond to dynamic incentives while controlling for differences in expected health-

[27] This value was obtained by regressing $\bar{P}^e_{1,y+1}$ for our samples on the deductible amount. There was one observation per year.

[28] We use the estimated average spending on the first day of 2015 as the baseline to calculate these percentage reductions.

Table 4: Difference in reaction to dynamic incentives across groups (controlling for risk score)

	effect of dynamic incentives (γ^e)		range of $\bar{P}^e_{1,y+1}$		utilization at beginning of year[a]		average N	effect of €100 increase in deductible[b]		
								change in $\bar{P}^e_{1,y+1}$ (in pp)	% reduction in utilization[c]	
	daily expenditure	any daily expenditure	min.	max.	daily expenditure	any daily expenditure			daily expenditure	any daily expenditure
baseline sample	-10.762	-0.0938	0.1443	0.2200	12.211	0.0733	27229	3.4	-3.0 %	-4.4 %
	(1.488)	(0.023)								
female	-9.232	-0.0781	0.1230	0.2057	11.687	0.0778	15503	3.7	-3.0 %	-3.7 %
	(2.023)	(0.027)								
male	-15.147	-0.0959	0.1718	0.2400	13.036	0.0680	11726	3.0	-3.5 %	-4.3 %
	(4.435)	(0.026)								
below median income[d]	-12.341	-0.1171	0.1396	0.2121	13.593	0.0839	11456	2.7	-2.5 %	-3.8 %
	(2.823)	(0.034)								
above median income	-9.827	-0.0786	0.1485	0.2375	11.401	0.0641	14786	3.7	-3.2 %	-4.5 %
	(2.983)	(0.022)								
age below 45	-3.020	-0.0274	0.2408	0.3442	9.320	0.0423	8272	4.2	-1.4 %	-2.8 %
	(3.822)	(0.019)								
age 45 and above	-12.103	-0.0911	0.0976	0.1829	13.040	0.0818	18957	3.8	-3.5 %	-4.2 %
	(1.518)	(0.027)								

Notes: This table shows the effects of dynamic incentives by subgroup. For this, the discontinuities and $\bar{P}^e_{1,y+1}$ were estimated within subsamples of our data. When doing so, we re-weight observations so that the distribution of the risk score in each re-weighted subsample is equal to the distribution in the baseline sample. To obtain the effect of a 10 percentage point change in $\bar{P}^e_{1,y+1}$ on expenditures, the reported estimates have to be divided by 10. The third and fourth column report the domain of $\bar{P}^e_{1,y+1}$ for each of the subgroups. The last two columns quantify the effect of increasing the deductible by €100 on healthcare utilization for the first day of the new year. We only look at utilization on the first day as all individuals are, by definition, below the deductible.

[a] Daily utilization at the beginning of the year was estimated using a local linear regression on the first regular day of 2015 (see Section 4). The observations are weighted such that the distribution of the riskscore in each re-weighted sample is equal to the distribution of the riskscore in the baseline sample.

[b] The effect of an increase in the deductible by €100 is only for the first day of 2015. For the effect of such an increase in the deductible on expenditures throughout the year, refer to Section 8.

[c] The estimated average utilization (not pseudo-censored) on the first day of 2015 was used as the base for these percentage change computations.

[d] Income is average income at the 6-digit postal code level. Income data were missing for around 1000 individuals across all year-pairs.

Table 5: Difference in reaction to dynamic incentives across sickness measure

	effect of dynamic incentives (γ^e)		range of $\bar{P}^e_{1,y+1}$		utilization at beginning of year[a]		average N	effect of €100 increase in deductible[b]: change in $\bar{P}^e_{1,y+1}$ (in pp)	effect of €100 increase in deductible[b]: % reduction in utilization[c]	
	daily expenditure	any daily expenditure	min.	max.	daily expenditure	any daily expenditure			daily expenditure	any daily expenditure
baseline sample	-10.762	-0.0938	0.1443	0.2200	12.211	0.0733	27229	3.4	-3.0 %	-4.4 %
	(1.488)	(0.023)								
riskscore										
above median	-17.563	-0.1582	0.0945	0.1482	13.393	0.0844	22172	2.2	-2.9 %	-4.1 %
	(1.748)	(0.043)								
below median	-2.359	-0.0154	0.3022	0.5061	7.553	0.0302	5509	9.3	-2.9 %	-4.7 %
	(0.727)	(0.011)								
percentile of spenders										
top 5 percentile	-47.959	-0.1375	0.0557	0.1103	22.016	0.1146	2437	1.6	-3.6 %	-2.0 %
	(21.925)	(0.113)								
top 10 percentile	-39.686	-0.1880	0.0954	0.1498	19.838	0.1050	5173	1.7	-3.3 %	-3.0 %
	(9.604)	(0.082)								
top 20 percentile	-24.028	-0.1574	0.1106	0.1711	17.043	0.0951	11728	2.3	-3.2 %	-3.8 %
	(5.053)	(0.056)								

Notes: This table shows the effects of dynamic incentives by sickness measure. For this, the discontinuities and $\bar{P}^e_{1,y+1}$ were estimated within each subsample. To obtain the effect of a 10 percentage point change in $\bar{P}^e_{1,y+1}$ on expenditures, the reported estimates have to be divided by 10. The third and fourth column report the domain of $\bar{P}^e_{1,y+1}$ for each of the subgroups. The last two columns quantify the effect of increasing the deductible by €100 on healthcare utilization for the first day of the new year. We only look at utilization on the first day as all individuals are, by definition, below the deductible.

[a] Daily utilization at the beginning of the year was estimated using a local linear regression on the first regular day of 2015 (see Section 4). The observations are weighted such that the distribution of the riskscore in each re-weighted sample is equal to the distribution of the riskscore in the baseline sample.

[b] The effect of an increase in the deductible by €100 is only for the first day of 2015. For the effect of such an increase in the deductible on expenditures throughout the year, refer to Section 8.

[c] The estimated average utilization (not pseudo-censored) on the first day of 2015 was used as the base for these percentage change computations.

[d] Income is average income at the 6-digit postal code level. Income data were missing for around 1000 individuals across all year-pairs.

care needs as measured by risk scores.[29] For this, we split the sample, alternatively according to gender, average income in the neighborhood below and above median, and ages below 45 years and 45 years and older. For each subsample, we weigh observations according to the inverse of the probability that the risk score falls into 5 bins formed by risk score quintiles. This means that for our weighted data, the distribution of risk scores is (approximately) the same as for the baseline sample.[30]

Column 1 of Table 4 shows our estimates of γ^e for different subgroups in the population. We find that almost every subgroup exhibits forward-looking behavior. The estimates of γ^e are significantly different from zero for all subgroups except for individuals that are below age 45. Point estimates are similar for males and females, and for individuals living in high and low income neighborhoods.

The results for any daily expenditure, shown in Column 2, show a similar pattern. Table 4 also shows the range of $\bar{P}^e_{1,y+1}$ across years, and utilization at the beginning of the year for each subgroup. These numbers are weighted by risk score quintiles.

In addition, we quantify the reduction in utilization (in percentage terms) on the first day of a new year if the deductible were to be increased by €100; changes in $\bar{P}^e_{1,y+1}$ are shown in the third-to-last column of Table 4, and changes in healthcare consumption (in percentage terms) are presented in the last two columns of Table 4. An increase in the size of the deductible by €100 leads to a -1.5% change in daily expenditures for individuals below age 45. For the remaining groups, we find that an increase in the size of the deductible by €100 leads to a percentage change in daily expenditures that ranges from -2.5% to -3.5%.

Our finding that the response to deductibles does not vary by income are in line with results from the RAND health insurance experiment (Manning et al., 1987) and with results for the Netherlands by Hayen et al. (2021). However, Hayen et al. (2021) find different responses by gender, but not by age. An explanation for the different findings on heterogeneous effects by gender and age could be that Hayen et al. (2021) estimate the response to static incentives while this study estimates the response to dynamic incentives.

In Table 5, we show heterogeneous effects of dynamic incentives by 2 measures of predicted sickness: (i) ex-ante riskscores and (ii) top spenders until the end of August of year y. Since we explore effects across these sickness measures, we do not weigh our data to make the distribution of risk scores comparable across subsamples (as we did before).

We first show that both patients with above median risk scores and patients with below median risk scores respond to dynamic incentives.[31] The estimate for γ^e is larger in absolute

[29] We use a linear regression to predict annual expenditures in year y based on age, gender, diagnosis for chronic conditions derived from pharmaceutical use, and medical spending in year $y-1$. The risk score of an individual is given by her predicted annual expenditures divided by average annual expenditures. The larger the riskscore, the more a person is predicted to spend, relative to the average. The same risk score measure is also used in Hayen et al. (2021).

[30] We then apply the same estimation approach, using the constructed weights, as for the baseline sample.

[31] We use the estimates for γ^e in Table 5 for patients with above and below median risk scores also in Section 8 when we compute the effect of an increase in deductibles on annual healthcare spending.

terms for patients with above median risk-scores than for patients with below median risk scores (-17.56 versus -2.36). However, the relative reduction in spending at the beginning of year $y+1$ as a result of a €100 increase in the deductible is similar for both groups. This holds for both total spending and the probability of any claim.

Table 5 also shows results for the top 5, 10, and 20 percent of spenders until the end of August of year y. We find that these groups strongly respond to dynamic incentives. Although the absolute value of the γ^e estimate is larger for sicker patients, the relative reductions in spending as a result of a €100 increase in the deductible are overall quite similar to the reductions for the baseline sample. Another finding in Table 5 is that even patients with very high previous spending experience changes in $\bar{P}^e_{1,y+1}$ across years. For example, the $\bar{P}^e_{1,y+1}$ for the top 5 percent of spenders up to August of year y ranges from 0.0557 to 0.1103. For comparison, the $\bar{P}^e_{1,y+1}$ for the baseline sample ranges between 0.1443 and 0.220 across year-pairs. Thus, if we compare the the top 5 percent of spenders and the baseline sample, the level of $\bar{P}^e_{1,y+1}$ is very different, but the difference between the highest and lowest $\bar{P}^e_{1,y+1}$ during our study period is comparable.

7 Robustness

Our main identifying assumption is that if there are no changes in dynamic incentives, then changes in healthcare utilization around the turn of the year should (in expectation) be constant across year-pairs. This assumption is formally stated in (2) in Section 3.3, and it allows us to attribute the negative relationship between the changes in healthcare consumption around the turn of the year and $\bar{P}^e_{1,y+1}$ to dynamic incentives. Thus, threats to our empirical strategy stem from any other possible reason that could explain such a negative relationship. In this section, we first discuss potential alternative explanations for our findings in Section 7.1, then we examine the effect of changes in dynamic incentives on types of care that are exempt from deductibles in Section 7.2, and finally we present alternative empirical specifications in Section 7.3.

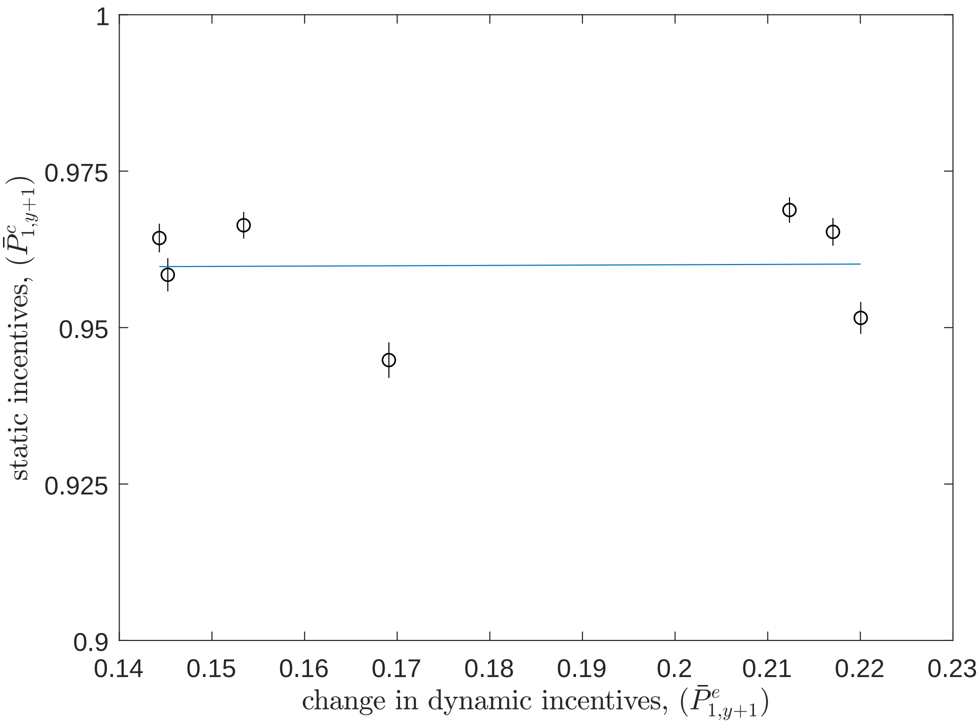
7.1 Alternative explanations

In the following, we discuss a wide range of potential alternative explanations for our findings such as changes in static incentives, salience of deductible amounts, endogenous selection into our sample based on forward-looking behavior, strategic timing of medical care use, changes in macroeconomic conditions, and changes in price levels and insurance coverage.

7.1.1 Changes in static incentives: relationship between $\bar{P}^e_{1,y+1}$ and $\bar{P}^c_{1,y+1}$

An increase in the deductible can potentially increase both current and future prices at the beginning of $y+1$. Our empirical strategy attributes differences in discontinuity sizes across year-pairs to differences in $\bar{P}^e_{1,y+1}$. One concern could be that the negative relationship between

Figure 5: Current price changes



Notes: This figure replicates Figure 4 for the current price as the outcome.

changes in healthcare consumption around the turn of the year and $\bar{P}^e_{1,y+1}$ shown in Figure 4 can be explained by changes in the average current price, $\bar{P}^c_{1,y+1}$, instead of changes in $\bar{P}^e_{1,y+1}$. Even though we look at days early in January, it is possible that some individuals exceed the deductible already within the first days of the year. Given that it requires more spending to cross the deductible (and thus have a current price of 0) in years with a higher deductible, it could be that $\bar{P}^c_{1,y+1}$ is higher in years with a higher deductible. This would result in a positive correlation between $\bar{P}^c_{1,y+1}$ and $\bar{P}^e_{1,y+1}$. In this case, we would falsely ascribe the effects of static incentives to the effects of dynamic incentives.

We can test whether $\bar{P}^e_{1,y+1}$ and $\bar{P}^c_{1,y+1}$ are indeed positively correlated. For this, we compute $\bar{P}^c_{1,y+1}$, the share of individuals who have not exceeded the deductible limit by $t = 1$ in year $y+1$, for all year-pairs in our data and plot them against $\bar{P}^e_{1,y+1}$. Figure 5 shows that there is little variation in $\bar{P}^c_{1,y+1}$ across years. The variation in $\bar{P}^c_{1,y+1}$ is not systematically related to $\bar{P}^e_{1,y+1}$, and the slope coefficient of the corresponding regression line, shown in the first row of Table 6, is close to 0 and not statistically significant. Thus, we conclude that our results cannot be attributed to changes in current prices instead of expected end-of-year prices.

Table 6: Robustness

Section	specification	slope		MR test	Up test	Down test
7.1.1	current price	0.005	(0.1131)	1.000	0.000	0.000
7.1.2	indicator for any spending below €155	-0.079	(0.0193)	0.946	0.000	0.290
7.1.3	indicator for any spending above €500	-0.004	(0.0029)	0.214	0.060	0.516
7.1.4	condition on large spending event in y	-12.884	(1.8245)	0.151	0.057	0.953
7.1.6	condition on age at least 67	-33.758	(5.9674)	0.538	0.023	0.600
7.1.7	inflation corrected expenditure (Eurostat)	-8.628	(2.9630)	0.189	0.002	0.958
7.1.8	inflation corrected expenditure (own deflator)	-10.254	(3.0674)	0.136	0.000	0.970
7.2.1	spending GP care	0.323	(0.6303)	1.000	0.169	0.000
7.2.2	spending individuals of age under 18	1.215	(2.3187)	0.225	0.729	0.508

Notes: This table replicates Table 3 for alternative outcomes and samples. See text in section in the first column for details, respectively. See also Table A.4 in the Online Appendix for additional results at the extensive margin.

7.1.2 Changes in static incentives: higher deductible payments

Higher deductible amounts can lead to higher payments. This can affect care consumption and can be seen as the effect of static incentives. At the same time it may not affect current prices at the beginning of $y+1$. To see this, consider a treatment of €300. At the beginning of the year 2009, with a deductible of €155 a patient has to pay €155 out-of-pocket for this treatment and the current price (for the last unit of care) is 0 if she receives it. In the year 2015, with a deductible of €375 a patient would have to pay €300 out-of-pocket, and the current price is would only be 1 if she received the treatment. But if the patient would decide in favor of the treatment in 2009 and against the treatment in 2015, then the current price would be zero in both years (in line with the empirical evidence we present above), but static incentives would be stronger in 2015 and would still have led to lower care consumption at the beginning of 2015. If such spending events are common, then they could provide an alternative explanation for the negative relationship between changes in healthcare consumption around the turn of the year and $\bar{P}^e_{1,y+1}$.

In order to address this concern, we first note that daily claims above €155 (the lowest deductible in our study period) are relatively rare, even among days with positive spending. Figure 2 shows that €155 is at the 86th percentile of the distribution of daily spending (conditional on any spending). In addition, we repeat our analysis for claims that are smaller than €155. Such claims are below the lowest deductible amount during our study period, which means such spending will lead to a current price of 1 in all years and patients will have to pay all costs out-of-pocket.[32] We define a binary variable that is 1 if daily spending is less than €155 (but greater than 0), and we repeat our analysis with this new outcome variable. The coefficient shown in row 2 of Table 6 shows that 1) spending that is always lower than the deductible amount does respond to changes in $\bar{P}^e_{1,y+1}$, and 2) the size of the response is comparable to the size of the

[32]This is true if there was no other previous spending in year $y+1$.

extensive margin response for our baseline results shown in Table 3.[33]

From this result, coupled with the result from the preceding section, we conclude that our results cannot be attributed to changes in static incentives that come from higher deductibles in later years.

7.1.3 Salience of deductible amounts

Another concern is that patients are not forward-looking, but that some features of the insurance contract such as deductible amounts are very salient. Thus, patients react in fact to deductible amounts and not to dynamic incentives. Salience predicts that patients respond to increasing deductibles even in situations where there is no variation in $\bar{P}^e_{1,y+1}$, and according to theory they should not respond.

One such situation would be for patients who cross the deductible limit with certainty. However, as Table 5 shows, even for patients with very high previous spending, there is still substantial variation in $\bar{P}^e_{1,y+1}$, and patients respond to this variation. Therefore, we cannot use this to test whether our results can be explained by salience.

Another situation where patients should not respond to changing deductible amounts is for large spending events above the maximum deductible amount in our study period. In our model, prices always refer to the last unit of care, and for large spending events both the current price and the expected end-of-year price are always 0 in all years. At the same time, if patients are spending less in years with a higher deductible because of the salience of deductible amounts, then we would expect that also high spending events occur less often in those years. Therefore, we can conduct a placebo test by examining whether the probability of spending more than €500 is affected by changes in dynamic incentives. The results in Table 6 show that the estimated coefficient for $\bar{P}^e_{1,y+1}$ is close to 0 and not statistically significant. Results for spending events larger than €5000 in Table A.4 in the Online Appendix are also in line with the predictions of our model. Thus, these placebo tests suggest that a simple version of the salience theory cannot explain our results. At the same time, we cannot rule out that more refined versions of the salience theory at least partly explain our results. However, these theories would need to assume that patients only react to the size of the deductible in certain situations.

7.1.4 Selection on forward-looking behavior

A further concern is endogenous sample selection. In all year-pairs we select our sample based on the top 38% of spenders by the end of August of year y. Yet, it is still possible that more forward-looking individuals may be sampled more frequently in later years with higher deductibles. To ensure that our results are not biased by this, we use an alternative sample selection procedure that selects our sample based on events that are not influenced by dynamic incentives. Above, we have shown that there is no effect of dynamic incentives on large spending events

[33] If we inflation-adjust the threshold of €155, our results are very similar.

above the maximum deductible amount, as implied by our model. We thus additionally restrict our sample (over and above the percentile-matching restriction) to individuals who exceed the deductible by August in year y with at least 1 daily spending event that is larger than €500.[34] This restriction reduces our sample size, for each year-pair, roughly by around half. Table 6 report estimation results for the restricted sample. The estimated effects are very similar to our baseline results shown in Table 3. This suggests that endogenous sample selection cannot explain our results.

7.1.5 Strategic timing of medical care use

Another threat to our empirical approach could be strategic timing of medical care use.[35] Since individuals in our estimation sample have crossed the deductible by August of year y, care at the beginning of year $y+1$ is always more expensive than at the end of year y. Thus, they have an incentive for strategic front-loading of medical care, i.e., shifting care from the beginning of year $y+1$ to the end of year y. If the amount of strategic front-loading increases with a higher $\bar{P}^e_{1,y+1}$ then this can potentially explain our estimation results.

In order to examine whether strategic front-loading is likely to affect our estimation results, we apply a test similar to one that is used by Brot-Goldberg et al. (2017).[36] The underlying rationale of our test is that if strategic front-loading occurs, then more care is shifted to months immediately before the end of the year than to months earlier in the year. Thus, if strategic front-loading increases with a higher $\bar{P}^e_{1,y+1}$, holding all other factors equal we would expect that spending in the months of November and December in year y is higher for year-pairs with a higher $\bar{P}^e_{1,y+1}$, relative to spending in October. We can test this hypothesis by running the following regression:

$$\bar{c}_{m,y} = \alpha_y + \lambda_m + \psi_m \bar{P}^e_{1,y+1} + \varepsilon_{m,y}, \tag{3}$$

where $\bar{c}_{m,y}$ is average (across individuals) monthly spending in month m and year y. α_y are year fixed effects, λ_m are month fixed effects, and $\bar{P}^e_{1,y+1}$ is the future price in year $y+1$. We implement the regression above by looking only at the last three months of the year, and we estimate two parameters for ψ_m, one for November and another one for December. These parameters ψ_m measure extra spending in November and December in year y for year-pairs with a higher $\bar{P}^e_{1,y+1}$, relative to spending in October. Thus, if strategic front-loading is present, we should expect the parameters ψ_m to be positive. We look at three different outcomes: average monthly spending, average pseudo-censored monthly spending and the probability of any spending at the monthly level. The estimated ψ_m, for each outcome variable, are reported in Table 7. The

[34] An alternative approach would have been to select a sample based on emergency care. However, our data do not provide a feasible way to identify such care.

[35] For instance, Cabral (2016) finds evidence for strategic delay of medical care use in the context of dental care.

[36] In our study, we interact ψ_m with $\bar{P}^e_{1,y+1}$ to obtain the main explanatory variables in (3). In their setting, Brot-Goldberg et al. (2017) can use the equivalent of ψ_m without interaction terms as main expnantory variables.

Table 7: Test for anticipatory spending

	avg. monthly spending	avg. monthly spending (PC)	prob. any monthly spending
ψ_{Nov}	77.98 (281.739)	−101.04 (59.842)	−0.2227 (0.1947)
ψ_{Dec}	−28.40 (322.591)	−70.43 (58.318)	−0.2828 (0.1734)

Notes: This table reports estimates of ψ_m based on (3). See text for details.

results indicate that there is no evidence for strategic front-loading. Also, the negative coefficients point to the opposite direction compared to what we would expect in the presence of strategic front-loading.

We can think of two possible explanations for these results. First, unlike for delaying medical care use, the scope for front-loading is likely to be limited since people typically need a medical indication before they can get treatment. Second, if there is a possibility to front-load then forward-looking patients should always do it. If there is the same amount of front-loading in all year-pairs $\{y, y+1\}$ then this does not bias our estimation results.[37]

7.1.6 Macroeconomic developments and analysis for persons of age 67+

A further alternative explanation could be that our results can be attributed not to changes in dynamic incentives but to changes in economic conditions. Purchasing power was increasing at the beginning and at the end of our study period, but it was decreasing between the years 2010 and 2013. Unemployment rates increased almost throughout the study period, from 3.7% in the year 2008 to 6.9% in the year 2014.[38]

If responses to cost-sharing become stronger with adverse economic conditions, then this could potentially explain the stronger decline in spending around the turn of the year for the later years in our study period. However, several pieces of evidence speak against this explanation. First, Table 4 shows that our estimated effect is not stronger for individuals in low-income neighborhoods than for individuals in high-income neighborhoods. Second, our estimated effect also is not stronger for persons below age 45 who were more affected by deteriorating labor market conditions during our study period than for individuals above age 45. Third, persons of age 67+ are for the most part retired and not affected by unemployment, and they received stable incomes from state pensions and sector-specific pension funds throughout our study period. In

[37] The above analysis also provides a test for another possible alternative explanation of our findings: mean reversion. Our estimation samples are selected based on the top 38% of spenders in year y. If high spending up to August in year y is followed by lower spending in later periods then this could explain lower spending at the beginning of year $y+1$. However, the above test shows that in the months of November and December of year y there is no significantly different spending in years with higher $\bar{P}^e_{1,y+1}$, other things being constant. This finding speaks against mean reversion as an explanation of our findings.

[38] Table A.2 shows annual unemployment rates and annual changes in purchasing power in the Netherlands during our study period.

row 5 of Table 6 we show that also persons of age 67+ strongly respond to dynamic incentives.[39]

Together, these findings indicate that changing macroeconomic conditions cannot explain our results. An underlying explanation could be that the social safety net dampened the effect of macroeconomic fluctuations on household income, and unemployment rates in the Netherlands during the great recession never reached the very high levels experienced in some other countries.

7.1.7 Adjustment for inflation

Costs of claims and the contents of the basic health insurance package are adjusted at the beginning of the year. These factors can have an influence on the size of discontinuities in medical spending around the turn of the year, and they can thus provide a potential alternative explanation for our results. In the baseline specification shown in Table 3 we use nominal prices which are not adjusted for inflation. In Table 6 we show that the slope coefficient is similar to the baseline specification if we adjust expenditures using Eurostat indices on healthcare prices in the Netherlands.[40]

7.1.8 Adjustment for sample-specific prices and changes in insurance coverage

A shortcoming of Eurostat price measures is that they are influenced by changes to prices above our censoring amount of €500 and by prices for services that do not count towards the deductible. They also do not capture changes of the services included in the basic package. In order to correct for these limitations of the Eurostat deflator, we create our own annual expenditure deflator using data for periods in each year where the price of healthcare is 0. Recall that individuals were selected into our sample conditional on having crossed the deductible by the end of August in the first year of the year-pair. This implies that these individuals face a 0 price of healthcare from September to December in that first year. We take the ratio of average pseudo-censored expenditures across these months, for a given year, to the average pseudo-censored expenditures for a base year. This produces an expenditure deflator that accounts for differences in expenditures due to changes in the cost of claims and the basic package, i.e., changes that are not influenced by changes in cost-sharing incentives.[41] We report the results from our empirical strategy after deflating expenditures in Table 6.[42] The resulting slope coefficient is almost identical to our baseline specification. Thus, we conclude that our results cannot be attributed to price changes or to changes in insurance coverage.

[39]The higher point estimate compared to the baseline can be attributed to higher healthcare spending levels for older persons. For this group, we could not adjust the sample weighting by risk score quartiles, as we have done in Table 4, because persons in this age group are never in the lowest risk-score quintile.

[40]Deflators using Eurostat numbers are reported in Table A.3 in the Online Appendix. Figure A.15 in the Online Appendix depicts the relationship between the estimated changes and $\bar{P}^e_{1,y+1}$ using deflated pseudo-censored expenditures based on inflation adjusted expenditures.

[41]The values for our expenditure deflator for each year are reported in the Online Appendix in Table A.3.

[42]We depict the relationship between the estimated changes and $\bar{P}^e_{1,y+1}$ in Figure A.17 in the Online Appendix.

7.2 Effects on types of care that are exempt from deductibles

In this section, we examine the effect of dynamic incentives on types of care that are exempt from deductibles.

7.2.1 Effect on GP care

In the Netherlands, care provided by GPs is exempt from deductible payments. Yet, there could still be an effect of dynamic incentives on GP care if patients perceive GP care either as a substitute or complement to care that falls under the deductible. GPs have a gatekeeper role in the Dutch healthcare system. Therefore, it is necessary to first visit a GP before seeing a specialist or obtaining a prescription. We run our analysis with the exact same specification as in the baseline and with spending on GP care as outcome variable. The result is reported in Table 6. We do not see a significant effect of $\bar{P}^e_{1,y+1}$ on GP care, and the coefficient is small compared to the estimates in the main analysis. Thus, we do not find evidence that GP care is either a complement or substitute to care that falls under the deductible.

7.2.2 Effect on minors under age 18

We also implement our empirical strategy for individuals under 18 years of age. Given that they do not face any cost-sharing, they should not exhibit the forward-looking behavior we document for our baseline sample of adults. The result is reported in Table 6. We indeed find no significant effect of dynamic incentives. In fact, in line with minors not facing any cost-sharing and the hypothesis that health care needs are comparable in December and January, none of the estimated discontinuities are significantly different from zero.

7.3 Alternative model specifications

Our findings are robust to a number of alternative specification of our empirical model, such as the use of an estimator that does not account for time trends, changes in the length of the "donut hole", the use of weekly instead of daily expenditure data, the inclusion of weekend care, and alternative measures of healthcare spending.

7.3.1 Not accounting for time trends

In our baseline model, we allow for time trends in healthcare spending in the periods before and after the turn of the year. This accounts for example for a gradual increase in spending as more patients exceed the deductible limit in the first days of year $y+1$. We also estimate a simpler specification with an uniform kernel that does not allow for time trends in spending (which implies our RD estimates are simple averages of spending before and after the turn of the year). Results are shown in rows 8 and 9 of Table A.4 in the Online Appendix, and they are similar to the baseline specification with linear time trends.

7.3.2 Changes in the donut hole

In our "donut hole" RD design we omit some days around the turn of the year due to holidays. For example, for the year-pair $\{2008, 2009\}$ we compare expenditures on Thursday December 18 and Thursday January 8. In robustness checks we change the length of the donut hole, and we compare averages across different weekdays. Specifically, we consider increasing the length of the donut hole by one week, and look at Wednesdays instead of Thursdays. Results are shown in rows 14 through 19 of Table A.4 in the Online Appendix. Our conclusions from Section 6 are robust to these different specifications.

7.3.3 Weekends, weekly level data, alternative cutoff point for pseudo censoring, and log specification

We also repeat our empirical analysis including care that is consumed on weekends, with weekly-level data on care consumed on weekends, a cut-off for pseudo-censored expenditures of €5000 instead of €500, and a specification where utilization is measured by the log of daily expenditures plus one. The results are presented in Table A.4 in the Online Appendix. Our conclusions do not change.

8 The effect of changes in deductibles on annual expenditures

In the previous sections, we have examined the relationship between the expected end-of-year price and daily care utilization around the turn of the year. In this section, we show how our estimate of γ^e can be used to predict the effect of a change in the size of the deductible on healthcare utilization at the annual level. Specifically, we quantify the euro amount saved, per capita in a year, from an increase in the size of the deductible by €100—from €375 to €475.

In a nutshell, we take the reduced-form relationship in (1) as the starting point. It relates care consumption at any point in time in a year to the current and the future price. The respective coefficients are given by γ^c and γ^e. For our simulations we use an estimate of γ^c from the literature and our estimate of γ^e. We use data for all years to estimate how average current and future prices at any point in time depend on the deductible size. Putting it all together gives an estimate of care consumption under the actual deductible and the counterfactual one. We can decompose the effect into a part stemming from static incentives and another part stemming from dynamic incentives.

8.1 Details and implementation

Recall that according to (1), care consumption is given by:

$$c_{it} = \kappa_{it} - \gamma^c \cdot P^c_{it} - \gamma^e \cdot P^e_{it}.$$

This holds at the individual level at any time t within a year. We take this as a starting point, aggregate over individuals, and make the dependence on the year explicit. y is our baseline year, 2015, with a €375 deductible. y' is a hypothetical year in which everything is the same, except that the size of the deductible is €475. We can write the resulting difference in average expenditures at time t as

$$\bar{c}_{t,y'} - \bar{c}_{t,y} = -\left[\gamma^c \cdot (\bar{P}^c_{t,y'} - \bar{P}^c_{t,y}) + \gamma^e \cdot (\bar{P}^c_{t,y'} \cdot \bar{P}^e_{t,y'|P^c_{it,y'}>0} - \bar{P}^c_{t,y} \cdot \bar{P}^e_{t,y|P^c_{it,y}>0})\right], \tag{4}$$

where $\bar{P}^e_{t,y|P^c_{it,y}>0}$ is the average expected end-of-year price in year y for individuals who have a positive current price at time t in year y. $\bar{P}^e_{t,y'|P^c_{it,y'}>0}$ is defined analogously.[43]

By writing (4) in this way, we make explicit that the effect of a change in the deductible consists of two parts. The first part is the change that is due to the change in static incentives. This is the product of $(\bar{P}^c_{t,y'} - \bar{P}^c_{t,y})$ and γ^c. $(\bar{P}^c_{t,y'} - \bar{P}^c_{t,y})$ is the change in the fraction of the population of individuals for whom the current price is 1 in period t of the year, when the deductible is increased by €100, and γ^c is the effect of the current price on care consumption.

The second part is the effect that is due to the change in dynamic incentives. It arises for all individuals who have a positive current price. The fraction of individuals for whom this is the case is given by $\bar{P}^c_{t,y}$ and $\bar{P}^c_{t,y'}$ in year y and y' respectively. These fractions are multiplied with the average expected end-of-year price conditional on a positive current price in each year. These quantities reflect the fact that when the deductible increases by €100, both the expected end-of-year price and the fraction of individuals who face this expected end-of-year price increases.

We can decompose the second part of (4) into two separate effects: (i) the effect due to the change in the fraction of individuals who face dynamic incentives (prevalence effect) and (ii) the effect due to the change in the expected end-of-year price (intensity effect) when the deductible is increased by €100:

$$\begin{aligned}
&\gamma^e \cdot (\bar{P}^c_{t,y'} \cdot \bar{P}^e_{t,y'|P^c_{it,y'}>0} - \bar{P}^c_{t,y} \cdot \bar{P}^e_{t,y|P^c_{it,y}>0}) \\
&= \gamma^e \cdot \left(\bar{P}^e_{t,y'|P^c_{it,y'}>0} \cdot (\bar{P}^c_{t,y'} - \bar{P}^c_{t,y}) + \bar{P}^c_{t,y} \cdot (\bar{P}^e_{t,y'|P^c_{it,y'}>0} - \bar{P}^e_{t,y|P^c_{it,y}>0})\right)
\end{aligned} \tag{5}$$

The first term in (5) relates to the prevalence effect, i.e., the increase in the share of individuals who face dynamic incentives when the deductible increases. This increased share of individuals is given by $(\bar{P}^c_{t,y'} - \bar{P}^c_{t,y})$ and is multiplied by the average expected end-of-year price

[43] Online Appendix A.4 establishes the link to the micro foundation of (1).

in year y'. The second term relates to the intensity effect, i.e., the increase in the expected end-of-year price for individuals who face dynamic incentives in year y. The share of individuals who face dynamic incentives in year y is given by $\bar{P}^c_{t,y}$ and this is multiplied by the change in the average expected end-of-year price when the deductible is increased by €100.

This forms the basis for predicting annual expenditures. For this, we use our estimate of γ^e. It measures the response to dynamic incentives . We have not estimated γ^c. Based on the literature, our starting point for the latter is that individuals reduce their expenditures by 40% if the current price increases from 0 to 1, conditional on the expected end-of-year price. 40% is very close to the estimates Brot-Goldberg et al. (2017) report for a sample of employees of a large firm in the U.S. and the estimates of Hayen et al. (2021) for the same data as we use in this paper. 40% is a relative effect. We also report results for a current price effect of 90%, 20%, and 0%, respectively. We translate these into absolute effects by multiplying them by average daily expenditures for individuals who have hit the deductible in 2015 (€15.7829).

We also need to predict prices. To that end, for each day t, we regress the share of individuals who have crossed the deductible and the expected end-of-year price (conditional on not yet having crossed) on the deductible amount in the corresponding year. Based on this, we predict all 4 prices in (4). The estimation equation is

$$outcome_{t,y} = \theta_{0,t} + \theta_{1,t} deductible_y + \eta_{t,y},$$

where $outcome_{t,y}$ is either $\bar{P}^c_{t,y}$ or $\bar{P}^e_{t,y|P^c_{it,y}>0}$. Here, we use the entire sample of individuals in our data from $2008-2015$. We account for the fact that individuals with higher riskscores react stronger to dynamic incentives by estimating γ^e separately for individuals above and below the median riskscore. We use the estimates presented in Table 5. We also account for the fact that these different groups may have different values of $\bar{P}^c_{t,y}$ and $\bar{P}^e_{t,y|P^c_{it,y}>0}$ by performing the underlying regressions for both groups in our sample separately. Figure A.4 in the Online Appendix shows our predictions of $\bar{P}^c_{t,y}$ and $\bar{P}^e_{t,y|P^c_{it,y}>0}$, for each riskscore group and deductible level. Aggregation over days in the year gives spending effects on yearly healthcare expenditure.

8.2 Results

Table 8 shows the results. Each row is for a different value of the current price effect. The first column shows the current price effect in percentage terms.[44] The second column shows the absolute effect on spending that is due to the change static incentives induced by increasing the deductible by €100. This is entirely due to patients facing a price of care of 1 for a longer time. The reduction in average annual expenditures that is due to current price changes is about €92

[44]The effects reported in the literature are all relative effects. To conduct the analysis we translate relative effects into absolute effects. These are denoted by γ^c in (1). This is done by multiplying the respective relative effect by the average spending of individuals who have crossed the deductible in 2015. The average is taken over individuals and days after the deductible was hit and is €18.26. The absolute effects are €16.43 for 90%, €7.30 for 40%, and €3.65 for 20%.

Table 8: Effect of increasing the deductible by €100 on annual expenditures

current price	reaction to	reaction to dynamic incentives		total
effect	static incentives	prevalence effect	intensity effect	effect
90%	206.21	87.09	33.31	326.61
40%	91.65	87.09	33.31	212.05
20%	45.83	87.09	33.31	166.22
0%	0	87.09	33.31	120.40

*Note*s: This table presents annual spending reductions under an assumed current price effect. See text for details.

when we calculate it for our baseline value of the current price effect of 40%.

Next, column 3 and 4 report the effect of dynamic incentives. We find that overall, healthcare expenditures decline by about €120 when the deductible size increases from €375 to €475. Given that, on average, an individual has total annual expenditures that count towards the deductible of around €2000 in 2015, this would imply increasing the deductible size by €100 to €475 leads to a 6% reduction in per capita healthcare expenditures due to changes in dynamic incentives.

The last column reports the total effect of dynamic and static incentives. For our baseline value of the current price effect of 40% we find that the reduction in average annual expenditures is €212, or 10.6%. Dynamic incentives account for more than half of the total reduction in annual spending.

9 Conclusion

In this paper, we show that a standard feature of deductible contracts, namely that they reset at the turn of the year, in combination with changes in deductible limits across years give rise to a differences-in-regression-discontinuities design that allows us to estimate the impact of changes in the size of the deductible on healthcare utilization at the beginning of the year.

Using administrative data from the Netherlands, we find that individuals are forward-looking and that the effect of dynamic incentives on healthcare expenditures is quantitatively important: a €100 increase in the deductible reduces our measure of daily expenditures at the beginning of the year, when individuals still have to pay for care themselves, by around 3.0% and reduces the probability of having any claim by around 4.4%.

We also explore whether the reaction to dynamic incentives differs across subgroups in our sample, defined by age, gender, and neighborhood income. Policy makers may be concerned if individuals with the same healthcare needs would react differently to dynamic incentives because they belong to different groups. Controlling for differences in healthcare needs we find that almost every subgroup exhibits forward-looking behavior. The only exception is the group of individuals that are below age 45. For the remaining groups, we find that an increase in the size of the deductible by €100 leads to percentage changes in daily expenditures that are similar to one another and range from -2.5% to -3.5%. In addition, we find effects that are similar to one another for subgroups defined by risk scores above and below the median and by high healthcare spending in previous periods, respectively. Also here, a €100 increase in the deductible size leads to a very similar percentage change in daily expenditures as for the baseline sample.

We interpret our findings through the lens of a model, which attributes these effects to changes in dynamic incentives. The main underlying idea is that static incentives are the same at the beginning of each year, as deductibles reset, while dynamic incentives change. Our micro-founded approach combines advantages of a model-based, structural approach and a reduced-form approach exploiting a natural experiment. In particular, we can use our estimates to predict the effect of dynamic incentives on annual expenditures. At the annual level, dynamic incentives imply that an increase in the deductible by €100 reduces annual healthcare expenditures (that count towards the deductible) by 6%. For the Netherlands, in the year 2015, this is equivalent to a €2 billion reduction in overall healthcare expenditures. This means that dynamic incentives have a first-order impact on healthcare utilization. In comparison, we predict that the response to static incentives reduces annual expenditures by 4.6% if the deductible is increased by €100. This prediction is based on an estimate of the effect of static incentives on healthcare expenditures from the literature. The relative size of the two effects suggests that patients' responses to dynamic incentives are an important part of the overall effect of cost-sharing schemes on healthcare expenditures—much more so than what the previous literature has suggested.

In a series of robustness checks we present evidence that speaks against alternative explanations. In particular, we show that differences in static incentives across years are unlikely to explain our results. We also show that our results cannot be explained by simple behavioral theories of salience that individuals demand less care when a particularly salient characteristic of their health insurance contract, the deductible amount, is higher. However, we cannot exclude that more refined behavioral theories of salience in which patients sometimes base their decisions on deductible amounts can at least partly explain our findings. This could well be the case, as we have shown in other work that patients do not only react to cost-sharing incentives, but that also the framing of these incentives is important (Hayen et al., 2021).

References

Abaluck, J., J. Gruber, and A. Swanson (2018). Prescription drug use under Medicare Part D: A linear model of nonlinear budget sets. *Journal of Public Economics 164*, 106–138.

Aron-Dine, A., L. Einav, A. Finkelstein, and M. Cullen (2015). Moral hazard in health insurance: Do dynamic incentives matter? *Review of Economics and Statistics 97*(4), 725–741.

Barreca, A. I., M. Guldi, J. M. Lindo, and G. R. Waddell (2011). Saving babies? revisiting the effect of very low birth weight classification. *Quarterly Journal of Economics 126*(4), 2117–2123.

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

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

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Cutler, D. and R. Zeckhauser (2000). The anatomy of health insurance. In J. Newhouse and A. Culyer (Eds.), *Handbook of Health Economics*, Volume 1, Chapter 11, pp. 563 – 643. Elsevier Science.

Dalton, C. M., G. Gowrisankaran, and R. J. Town (2020). Salience, Myopia, and Complex Dynamic Incentives: Evidence from Medicare Part D. *Review of Economic Studies 78*(2), 822–869.

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

Einav, L. and A. Finkelstein (2018). Moral hazard in health insurance: What we know and how we know it. *Journal of the European Economic Association 16*, 957 – 982.

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

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

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

Farbmacher, H., P. Ihle, I. Schubert, J. Winter, and A. Wuppermann (2017). Heterogeneous effects of a nonlinear price schedule for outpatient care. *Health Economics 26*(10), 1234–1248.

Gelman, A. and G. Imbens (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics 37*(3), 447–456.

Gerfin, M., B. Kaiser, and C. Schmid (2015). Healthcare demand in the presence of discrete price changes. *Health Economics 24*(9), 1164–1177.

Guo, A. and J. Zhang (2019). What to expect when you are expecting: Are health care consumers forward-looking? *Journal of Health Economics 67*, 102216.

Hanushek, E. (1973). Efficient estimators for regressing regression coefficients. *The American Statistician 28*, 66–67.

Hayen, A., T. Klein, and M. Salm (2021). Does the framing of patient cost-sharing incentives matter? The effects of deductibles vs. no-claim refunds. *Journal of Health Economics 80*(102520), 1–19.

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

Kaiser Family Foundation (2020). Employer health benefits: 2020 annual survey. Kaiser Family Foundation, San Francisco, CA, available at https://files.kff.org/attachment/Report-Employer-Health-Benefits-2020-Annual-Survey.pdf.

Keeler, E. B., J. P. Newhouse, and C. E. Phelps (1977). Deductibles and the demand for medical care services: The theory of a consumer facing a variable price schedule under uncertainty. *Econometrica 45*(3), 641–655.

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

Klein, T. J., M. Salm, and S. Upadhyay (2019). Proposition 1 by Ellis (1986) for a more general model with quasi-hyperbolic discounting and interest. Note, Tilburg University. Tilburg, The Netherlands.

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

McGuire, T. (2011). Demand for health insurance. In T. M. M.V. Pauly and P. Barros (Eds.), *Handbook of Health Economics*, Volume 2, Chapter 5, pp. 317 – 396. Amsterdam: Elsevier Science.

O'Donoghue, T. and M. Rabin (1999). Doing it now or later. *American Economic Review 89*(1), 103–124.

Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics*. Cambridge, United Kingdom: Cambridge University Press.

Patton, A. and A. Timmermann (2010). Monotonicity in asset returns: New tests with applications to the term structure, the capm, and portfolio sorts. *Journal of Financial Economics 98*, 605 – 625.

Online Appendix

A Micro foundation

In this appendix, we provide a micro foundation for our differences-in-regression-discontinuities analysis and the counterfactual simulations. For this, we propose an empirical model that predicts a reaction to both the current and the expected end-of year price and features quasi-hyperbolic discounting (Appendix A.1). We derive the implied reduced-form equation for care consumption at the individual level (Appendix A.2). Then, we show that our differences-in-regression-discontinuities estimates have a particular structural interpretation (Appendix A.3). Thereafter, we provide a micro-foundation for our counterfactual experiment that we conduct in Section 8 of the paper (Appendix A.4). We also provide details on the relationship of our model to various versions of the Keeler et al. (1977) model that have been used in the literature, with a particular emphasis on the question what constitutes a good measure of dynamic incentives and to what extent our measure is model-dependent (Appendix A.5). We end with a more general discussion (Appendix A.6).

A.1 Model

The year is divided into periods $t = 1, \dots, T$. In each period, patient i faces healthcare needs λ_{it} that are i.i.d. draws from a time-specific distribution $F_{\lambda t}$.

The model and the ensuing analysis are tailored to our setting where cost sharing is implemented using a standard deductible. At the beginning of each time period, i has a remaining deductible R_{it} and learns about her healthcare needs. She then chooses how much care c_{it} to consume. The remaining deductible is $R_{i1} = D$ at the beginning of the year and evolves according to $R_{it} = \max\{0, R_{i,t-1} - c_{i,t-1}\}$. Care consumption leads to out-of-pocket payments

$$C(c_{it}, R_{it}) = \min\{R_{it}, c_{it}\}. \tag{6}$$

It will be useful to define the out-of-pocket price for the last unit of care that is consumed in t as

$$P_{it}^{c} \equiv \frac{\partial C(c_{it}, R_{it})}{\partial c_{it}}.$$

Here, the superscript "c" stands for "current".

Flow utility is quasi-linear in money and given by

$$u(c_{it}; \lambda_{it}, R_{it}) = (c_{it} - \lambda_{it}) - \frac{1}{2\omega}(c_{it} - \lambda_{it})^2 - C(c_{it}, R_{it}). \tag{7}$$

This quadratic functional form has been used by Einav et al. (2013) in a different context—patients choosing which health insurance to buy—and can be seen as a quadratic approximation to any utility function that is defined on the difference between healthcare consumption c_{it} and healthcare needs λ_{it} and quasi-linear in money. One advantage of this specification is that the parameter ω is readily interpretable. To see this, it is useful to inspect the first order condition for an interior solution in the last period,[45]

$$1-\frac{1}{\omega}\cdot(c_{iT}-\lambda_{iT})-\frac{\partial C(c_{iT},R_{iT})}{\partial c_{iT}}=0.$$

This implies that the optimal static consumption choice in the last period is

$$c^*_{iT}=\lambda_{iT}+\omega\cdot\left(1-\frac{\partial C(c_{iT},R_{iT})}{\partial c_{iT}}\right).$$

The last term in parentheses is the marginal out-of-pocket cost for the last unit of care that is consumed in period T. In the case of a standard deductible that we study in this paper, this cost is either 1 by the end of T the patient will have exceeded the deductible limit, or 0. So, care consumption is given by

$$c^*_{iT}=\lambda_{iT}+\omega$$

for patients above the deductible limit and

$$c^*_{iT}=\lambda_{iT}$$

for patients below the deductible limit. This means that ω is the additional care consumption when individuals do not have to pay out-of-pocket for the last unit of care they consume.

In any earlier period t patients solve a dynamic decision problem when they choose c_{it}. The associated value function is

$$V_t(\lambda_{it},R_{it})=\max_{c_{it}} u(c_{it};\lambda_{it},R_{it})+\beta\delta\cdot\mathbb{E}\left[\tilde{V}_{t+1}(\lambda_{i,t+1},R_{i,t+1})\right], \tag{8}$$

$$\text{with } \tilde{V}_t(\lambda_{it},R_{it})=\max_{c_{it}} u(c_{it};\lambda_{it},R_{it})+\delta\cdot\mathbb{E}\left[\tilde{V}_{t+1}(\lambda_{i,t+1},R_{i,t+1})\right]. \tag{9}$$

The distinction between $V_t(\lambda_{it},R_{it})$ and $\tilde{V}_t(\lambda_{it},R_{it})$ arises because of quasi-hyperbolic discounting with naive individuals: if $\beta<1$ then they are too optimistic and wrongly expect that from the next period onward they will not suffer from present bias and only be exponential discounters.

The model outlined above is similar to the original theoretical model by Keeler et al. (1977). However, there are two key conceptual differences. First, patients can't save or borrow against future income. As we discuss in Section A.5.1, this is the most straightforward and internally consistent way to build a model in which patients react to the current price. The second conceptual difference is a generalization: patients are quasi-hyperbolic discounters. They discount all

[45] Keep in mind that the budget set is nonlinear. This means that the first order condition could hold at two values for c_{iT}. Also, notice that $C(c_{it},R_{it})$ is not differentiable at $c_{it}=R_{it}$. However, (6) implies that it will in general not be optimal to choose the kink point. For this reason, we will abstract from this in the following for the ease of the exposition (but would have pointed it out in the relevant places if it would have led to different conclusions). In the context of Medicare Part D, as is well-understood, one would have to instead carefully take bunching at the kink into account, as has been done for instance by Abaluck et al. (2018).

future utilities by a factor β and, in addition, utility that is τ periods in the future by δ^{τ}. Related to the discounting, we assume that individuals are naive in the sense of O'Donoghue and Rabin (1999), meaning that they do not foresee that their time preference will change in the future.

Our model is similar to the empirical models in Einav et al. (2015) and Abaluck et al. (2018). Also, these models do not feature savings or borrowing. However, the former assumes $\beta = 1$ and the latter $\delta = 1$. We provide additional formal results for $\beta < 1$ and general values of δ.

We now derive the optimal policy in the form of a reduced-form equation that relates today's healthcare consumption c_{it} to the state variables λ_{it} and R_{it}, and to beliefs about the future.

These beliefs are directly related to the probability of crossing the deductible in future periods. It will be useful to define

$$q_{it}(\tau) \equiv \Pr(R_{i\tau} > 0, c_{i\tau} > R_{i\tau} | R_{it}, c_{it}), \text{ where } \tau > t.$$

In words, this is the probability of exceeding the cost sharing limit in period τ, which refers to the joint event that the remaining deductible at the beginning of period τ is positive and that care consumption in period τ exceeds the remaining deductible. This probability is conditional on information available in t, so it is from the perspective of period t. In particular, we condition on the deductible at the beginning of period t and spending in t. This implies a value for the remaining deductible at the end of period t, which is the key state variable here. Notice that, in line with rational expectations over future healthcare needs and behavior in the future, this probability $q_{it}(\tau)$ can be estimated from data.

With this probability $q_{it}(\tau)$, a consumer who is spending one additional unit of money out-of-pocket in t will spend one unit of money less in τ. As utility is quasi-linear in money, the utility cost of this additional unit of care will then be $1 - \beta\delta^{\tau-t}$. 1 is the cost in the current period when utility is quasi-linear in money and $\beta\delta^{\tau-t}$ are the discounted savings in τ. These discounted savings refer to the chance that higher expenditures in period t might lead to lower expenditures in period τ. If it is uncertain when the individual will exceed the deductible, then the patient has to form expectations about this. The expected utility cost of one additional unit of care, will then be $P_{it} = 1 - \beta \cdot \sum_{\tau=t+1}^{T} \delta^{\tau-t} q_{it}(\tau)$. P_{it} depends on individual preference parameters β and δ, and it involves beliefs, as it depends on future realization of healthcare needs. It is the correct measure of dynamic incentives in this context.

Furthermore, $P_{it} = 0$ whenever $P_{it}^{c} = 0$. The expected utility cost of one additional unit of care consumption is zero whenever a patient has already exceeded the deductible limit before the end of year t. Therefore, we can write P_{it} as

$$P_{it} = \begin{cases} 0 & \text{if } P_{it}^{c} = 0 \\ 1 - \beta \cdot \sum_{\tau=t+1}^{T} \delta^{\tau-t} q_{it}(\tau) & \text{if } P_{it}^{c} = 1. \end{cases}$$

A.2 Reduced-form equation for care consumption

We now derive a reduced-form equation for care consumption at the individual level. We summarize our result in the following proposition. The proof resembles the proof by Ellis (1986), as discussed in Appendix A.5 below, and the one by Abaluck et al. (2018) who however impose $\delta = 1$.

Proposition 1. *In the model described in Appendix A.1, optimal care consumption is given by*

$$c_{it} = \lambda_{it} + \omega \cdot (1 - P_{it}), \tag{10}$$

where the relevant price is given by

$$P_{it} = \begin{cases} 0 & \text{if } P^c_{it} = 0 \\ 1 - \beta \cdot \sum_{\tau=t+1}^{T} \delta^{\tau-t} q_{it}(\tau) & \text{if } P^c_{it} = 1. \end{cases} \tag{11}$$

Proof of Proposition 1. The first order condition for the maximization problem in (8) is

$$\frac{\partial u(c_{it}; \lambda_{it}, R_{it})}{\partial c_{it}} + \beta\delta \cdot \frac{\partial \mathbb{E}\left[\tilde{V}_{t+1}(\lambda_{i,t+1}, R_{i,t+1})\right]}{\partial R_{i,t+1}} \cdot \frac{\partial R_{1,t+1}}{\partial c_{it}} = 0. \tag{12}$$

First consider the case when $R_{it} = 0$. In this case, patients solve a static problem, and the second term on the left hand side of (12) is zero. Hence, the first order condition in any period for which $R_{it} = 0$ is given by the derivative of the flow utility in (7) with respect to c_{it}, evaluated at $R_{it} = 0$, being equal to zero,

$$1 - \frac{1}{\omega} \cdot (c_{it} - \lambda_{it}) = 0,$$

which implies $c_{it} = \lambda_{it} + \omega$ for $R_{it} = 0$. By (11) we have that $P_{it} = 0$. So, (10) holds for all t whenever $R_{it} = 0$.

Next consider the case $R_{it} > 0$. Our goal is to show that

$$\frac{\partial \mathbb{E}\left[\tilde{V}_s(\lambda_{i,s}, R_{is})\right]}{\partial R_{is}} = -\sum_{\tau=s}^{T} \delta^{\tau-s} q_{it}(\tau), \tag{13}$$

for any $t < s \leq T$. If (13) holds for $s = t+1$ then it can be shown that the first order condition in (12) implies that equation (10) in the proposition is true.

We show that (13) holds by induction. First, we show that it holds for $s = T$. In that period, patients reach the cost sharing limit with probability $q_{it}(T)$ from the perspective of period t. For patients who reach the cost-sharing limit in period T, a remaining deductible that is higher by one unit means that they spend that one unit more out-of-pocket. By the envelope theorem, the derivative of (9) with respect to the remaining deductible in T is

$$\frac{\partial \mathbb{E}\left[\tilde{V}_T(\lambda_{i,T}, R_{iT})\right]}{\partial R_{iT}} = (-1) \cdot q_{it}(T) + 0 \cdot (1 - q_{it}(T)).$$

So, (13) holds for $s = T$.

It remains to show that (13) holds for s whenever it holds for $s+1$. By the envelope theorem, we have that the derivative of (9) with respect to R_{is} is

$$\begin{aligned}\frac{\partial \mathbb{E}\left[\tilde{V}_s(\lambda_{is}, R_{is})\right]}{\partial R_{is}} &= \frac{\partial u(c_{is}; \lambda_{is}, R_{is})}{\partial R_{is}} + \delta \cdot \frac{\partial \mathbb{E}\left[\tilde{V}_{s+1}(\lambda_{i,s+1}, R_{i,s+1})\right]}{\partial R_{i,s+1}} \\ &= -\frac{\partial C(c_{is}, R_{is})}{\partial R_{is}} + \delta \cdot \frac{\partial \mathbb{E}\left[\tilde{V}_{s+1}(\lambda_{i,s+1}, R_{i,s+1})\right]}{\partial R_{i,s+1}}.\end{aligned}$$

We have that

$$-\frac{\partial C(c_{is}, R_{is})}{\partial R_{is}} = \begin{cases} -1 & \text{with probability } q_{it}(s) \\ 0 & \text{with probability } 1 - q_{it}(s). \end{cases}$$

Using this and substituting in (13) gives

$$\frac{\partial \mathbb{E}\left[\tilde{V}_s(\lambda_{is}, R_{is})\right]}{\partial R_{is}} = (-1) \cdot q_{it}(s) + \delta \cdot \sum_{\tau=s+1}^{T} \delta^{\tau-(s+1)} \cdot (-1) \cdot q_{it}(\tau) = -\sum_{\tau=s}^{T} \delta^{\tau-s} q_{it}(\tau).$$

This completes the proof. □

Before we discuss how equation (10) in the proposition relates to our analysis we write it as

$$\begin{aligned} c_{it} &= \lambda_{it} + \omega \cdot \left[1 - P_{it}^c \cdot \left(1 - \beta \cdot \sum_{\tau=t+1}^{T} \delta^{\tau-t} q_{it}(\tau)\right)\right] \\ &= \lambda_{it} + \omega \cdot \left[1 - P_{it}^c \cdot \left(1 - \beta \cdot \left(\sum_{\tau=t+1}^{T} \delta^{\tau-t} q_{it}(\tau) - 1 + 1\right)\right)\right] \\ &= \lambda_{it} + \omega \cdot \left[1 - P_{it}^c \cdot \left(1 - \beta \cdot 1 - \beta \cdot \left(\sum_{\tau=t+1}^{T} \delta^{\tau-t} q_{it}(\tau) - 1\right)\right)\right] \\ &= \lambda_{it} + \omega \cdot \left[1 - \left((1-\beta) \cdot P_{it}^c + \beta \cdot P_{it}^c \cdot \left(1 - \sum_{\tau=t+1}^{T} \delta^{\tau-t} q_{it}(\tau)\right)\right)\right] \end{aligned}$$

where the first equality follows from substituting in the expression for P_{it}, the second from adding and subtracting 1 from the expression in the innermost set of parentheses, the third and fourth from rearranging terms.

As a next step, we show that this way to write the optimal policy relates care consumption c_{it} to the current price P_{it}^c and the expected end-of-year price P_{it}^e. To show this, we assume $\delta = 1$. In the main text, we have defined the expected end of year price as the probability that the patient will have to pay for the last unit of care in the year. Formally, we can write

$$P_{it}^{e} = P_{it}^{c} \cdot \left(1 - \sum_{\tau=t+1}^{T} q_{it}(\tau)\right).$$

If $P_{it}^{c} = 0$, then the expected end-of-year price is 0; if $P_{it}^{c} = 1$, then the expected end of year price is 1 minus the probability to hit the deductible limit in any future period. So, we have

$$c_{it} = \lambda_{it} + \omega \cdot [1 - ((1-\beta) \cdot P_{it}^{c} + \beta \cdot P_{it}^{e})]. \tag{14}$$

This way to re-write the optimal policy is useful because it forms a basis for a regression of care consumption on the current and the expected end-of-year price. Such a regression has been carried out for instance by Brot-Goldberg et al. (2017). Furthermore, this equation is equivalent to the first equation in the main text

$$c_{it} = \kappa_{it} - \gamma^{c} \cdot P_{it}^{c} - \gamma^{e} \cdot P_{it}^{e}.$$

where $\kappa_{it} \equiv \lambda_{it} + \omega$, the parameter on the current price is $\gamma^{c} \equiv \omega \cdot (1-\beta)$ and the parameter on the expected end-of-year price is $\gamma^{e} \equiv \omega \cdot \beta$.[46] The relative importance of γ^{c} and γ^{e} depends on the parameter for hyperbolic discounting β. If there is no hyperbolic discounting (β = 1), then patients will exclusively respond to the expected end of year price P_{it}^{e}. This is the case considered in the original paper by Keeler et al. (1977). If there is hyperbolic discounting ($\beta <$ 1) then patients will respond to both P_{it}^{e} and P_{it}^{c}.

Finally, observe that the relative size of the coefficients on the current and expected end-of-year price is informative about the extent to which individuals discount all future periods, as summarized by β. In particular, we have that

$$\frac{\gamma^{c}}{\gamma^{e}} = \frac{1-\beta}{\beta}.$$

or

$$\beta = \frac{1}{1 + \gamma^{c}/\gamma^{e}}.$$

This means that if γ^{c} and γ^{e} are identified, then also β is identified.

A.3 Interpretation of differences-in-regression-discontinuities estimates

In the main part of the paper, we relate changes in care consumption at the turn of the year to the expected end-of-year price. Here we provide a micro foundation. We do so under the assumption that $\delta = 1$. This allows us to follow the literature and measure changes in dynamic incentives by changes in the expected end-of-year price. In Appendix A.5 we further discuss the assumption that $\delta = 1$, and we show that even if we depart from this assumption P_{it}^{e} is closely related to the correct measure of dynamic incentives.

[46]This result for $\delta = 1$ is not new. Abaluck et al. (2018, p.110) provide a similar expression.

Denote averages taken over individuals in a given period t of year y by a bar indexed by t and y. Using this notation, we have that the average current price in $t=1$ of year $y+1$ is $\bar{P}^c_{1,y+1}$ and the expected end-of-year price at the beginning of the year is $\bar{P}^e_{1,y+1}$.[47]

Based on (14) we can write average care consumption in $t=1$ of year $y+1$ as

$$\bar{c}_{1,y+1} = \bar{\lambda}_{1,y+1} + \omega \cdot \left[1 - \left((1-\beta)\cdot \bar{P}^c_{1,y+1} + \beta \cdot \bar{P}^e_{1,y+1}\right)\right].$$

For each year-pair we use a sample of individuals whose price is zero by the end of year y. In terms of our model average care consumption in the last period of year y is thus

$$\bar{c}_{T,y} = \bar{\lambda}_{T,y} + \omega.$$

As discussed in Section (3), our identifying assumption in (2) is that the change in medical needs around the turn of the year is the same for all year-pairs. This assumption can also be expressed in terms of average medical needs $\bar{\lambda}_{t,y}$ instead of $\bar{\kappa}_{t,y}$. Then we have for all year-pairs $\{y, y+1\}$ and $\{y+1, y+2\}$

$$(\bar{\lambda}_{1,y+2} - \bar{\lambda}_{T,y+1}) = (\bar{\lambda}_{1,y+1} - \bar{\lambda}_{T,y})$$

We discuss the plausibility of this assumption in Section 7. Under this assumption, the change in the discontinuity from year-pair $\{y, y+1\}$ to, year-pair $\{y+1, y+2\}$ is

$$(\bar{c}_{1,y+2} - \bar{c}_{T,y+1}) - (\bar{c}_{1,y+1} - \bar{c}_{T,y}) = -\omega \cdot \left[(1-\beta)\cdot(\bar{P}^c_{1,y+2} - \bar{P}^c_{1,y+1}) + \beta \cdot (\bar{P}^e_{1,y+2} - \bar{P}^e_{1,y+1})\right]. \tag{15}$$

At the beginning of the year, the current price is one because the deductible resets at the turn of the year. At the end of the first period, it can be zero if an individual experiences a health shock in the first period and consumes more care than the deductible limit. However, in Section 7.1 we show that the difference $\bar{P}^c_{1,y+2} - \bar{P}^c_{1,y+1}$ does not significantly vary across years. Thus, our estimate for γ^e in (1) is an estimate of $\omega \cdot \beta$.

A.4 Micro-foundation for counterfactuals

In Section 8 we show how we can use an estimate of $\omega \cdot \beta$ to make a counterfactual prediction of healthcare expenditures for a different value of the deductible. We also show how one can use prior knowledge of the current price effect to say how much of the change in healthcare expenditures is driven by the current price effect and how much is driven by the reaction to dynamic incentives. In this section, we provide a micro foundation for this.

[47] Here we do not make a formal distinction between expectations and averages. In our analysis, we interpret estimates that are based on sample averages and are interested in the effect of a change in the deductible on average expenditures.

Maintaining the assumption that $\delta = 1$, it follows from (10) and (11) that

$$\begin{aligned}\bar{c}_{t,y} &= \bar{\lambda}_{t,y} + (1-\bar{P}^c_{t,y})\cdot\omega + \bar{P}^c_{t,y}\cdot\omega\cdot\left[1-\left(1-\beta\cdot(1-\bar{P}^e_{t,y|P^c_{it,y}>0})\right)\right]\\ &= \bar{\lambda}_{t,y} - \bar{P}^c_{t,y}\cdot\omega\cdot\left(1-\beta\cdot(1-\bar{P}^e_{t,y|P^c_{it,y}>0})\right),\end{aligned}$$

where $\bar{P}^e_{t,y|P^c_{it,y}>0}$ is the average expected end-of-year price from the perspective of period t, where the average is taken over individuals with a positive current price in period t of year y; analogously for y'. Now consider the case in which the only difference between year y and year y' is the size of the deductible. The difference in the expenditure between those two years is

$$\begin{aligned}\bar{c}_{t,y'} - \bar{c}_{t,y} &= -\left[\bar{P}^c_{t,y'}\cdot\omega\cdot\left(1-\beta\cdot(1-\bar{P}^e_{t,y'|P^c_{it,y'}>0})\right) - \bar{P}^c_{t,y}\cdot\omega\cdot\left(1-\beta\cdot(1-\bar{P}^e_{t,y|P^c_{it,y}>0})\right)\right]\\ &= -\left[(\bar{P}^c_{t,y'}-\bar{P}^c_{t,y})\cdot\omega\cdot(1-\beta) + \omega\cdot\beta\cdot(\bar{P}^c_{t,y'}\cdot\bar{P}^e_{t,y'|P^c_{it,y'}>0} - \bar{P}^c_{t,y}\cdot\bar{P}^e_{t,y|P^c_{it,y}>0})\right]\end{aligned}$$

This can also be written in a way that corresponds to the exposition in Section 8. For this let y' be the year with the higher deductible and denote by $\bar{P}^e_{t,y|P^c_{it,y'}>0}$ the average expected end-of-year price in period t of year y (alternatively y') for individuals who have a positive current price in period t of year y'. Using

$$\bar{P}^c_{t,y'}\cdot\bar{P}^e_{t,y|P^c_{it,y'}>0} = \bar{P}^c_{t,y}\cdot\bar{P}^e_{t,y|P^c_{it,y}>0},$$

we have

$$\bar{c}_{t,y'} - \bar{c}_{t,y} = -\left[(\bar{P}^c_{t,y'}-\bar{P}^c_{t,y})\cdot\omega\cdot(1-\beta) + \bar{P}^c_{t,y'}\cdot(\bar{P}^e_{t,y'|P^c_{it,y'}>0} - \bar{P}^e_{t,y|P^c_{it,y'}>0})\cdot\omega\cdot\beta\right].$$

We can see that the effect consists of two parts. $(\bar{P}^c_{t,y'} - \bar{P}^c_{t,y})$ is the change in the fraction of the population of individuals for whom the current price is 1. This is multiplied by the effect of the current price on care consumption, $\omega\cdot(1-\beta)$, and together gives the change that is due to the change in static incentives. The second part is the effect that is due to the change of dynamic incentives. It arises for all individuals who have a positive current price in year y', i.e. the fraction $\bar{P}^c_{t,y'}$, and is given by the change in the end-of-year price for those individuals times the reaction to the end-of-year price, $\omega\cdot\beta$.

A.5 Relevant measure of dynamic incentives across models

A.5.1 The model by Keeler et al. (1977) and the result by Ellis (1986)

We first describe a version of the Keeler et al. (1977)-Ellis (1986) model that is comparable to ours. The main generalization in our model is that patients are quasi-hyperbolic $\beta-\delta$ discounters.

Flow utility is additive in healthcare consumption and other consumption y_{it}. Instead of (7), we now have

$$u(c_{it}, y_{it}; \lambda_{it}) = (c_{it} - \lambda_{it}) - \frac{1}{2\omega}(c_{it} - \lambda_{it})^2 + u_c(y_{it}). \tag{16}$$

Notice that the costs of healthcare spending, unlike in (7), are not included in (16). These costs, instead, enter the optimization problem through the evolution of wealth across periods:

$$W_{t+1} = W_t - y_t - C(c_{it}, R_{it})$$

Allowing for quasi-hyperbolic discounters and interest, r, on wealth, we have the following Bellman equation:

$$\begin{aligned} V_{it}(\lambda_{it}, W_{it}) &= \max_{c_{it}, y_{it}} u(c_{it}, y_{it}; \lambda_{it}) + \beta\delta \cdot \mathbb{E}\left[\tilde{V}_{t+1}(\lambda_{i,t+1}, W_{i,t+1})\right] \\ \text{with } \tilde{V}_t(\lambda_{it}, W_{it}) &= \max_{c_{it}, y_{it}} u(c_{it}, y_{it}; \lambda_{it}) + \delta \cdot \mathbb{E}\left[\tilde{V}_{t+1}(\lambda_{i,t+1}, W_{i,t+1})\right] \end{aligned} \tag{17}$$

such that

$$\begin{aligned} W_{it+1} &= (1+r)(W_{it} - y_t - C(c_{it}, R_{it})) \\ R_{it+1} &= \max(R_{it} - c_{it}, 0) \\ V_{T+1} &= V(\lambda_{iT+1}) + V(W_{iT+1}) \end{aligned}$$

The main difference between (8) and (17) is that the costs of healthcare do not affect flow utility and only affects future values through its effect on W_{it+1}. An important related assumption, given by the terminal condition on V_{T+1}, is that there are no wealth effects in this formulation: the marginal utility of wealth in period $T+1$ is assumed to be equal to 1. With this, one can show that instead of (12) the first order condition is[48]

$$\frac{u(c_{it}, y_{it}; \lambda_{it})}{\partial c_{it}} - \beta \cdot \delta^{(T+1)-t} \cdot (1+r)^{(T+1)-t} \cdot \left(1 - \sum_{\tau=1}^{T} q_{it}(\tau)\right) = 0. \tag{18}$$

From this we get that optimal care consumption is given by

$$c_{it} = \lambda_{it} + \omega\left(1 - \beta\delta^{(T+1)-t}(1+r)^{(T+1)-t} \cdot \left(1 - \sum_{\tau=1}^{T} q_{it}(\tau)\right)\right).$$

The following proposition summarizes the above.

Proposition 2 (Relevant price in Ellis (1986) with discounting and interest)**.** *Assume the model*

[48] In a separate note, Klein et al. (2019), we derive this result using a more general version of the original setup and the same notation as in Ellis (1986). This note is not meant for publication, but available upon request from the authors.

setup in A.1, with the exception that: (i) patients are endowed with wealth and can freely save and borrow; (ii) utility is not quasi-linear in money, but given by (16); (iii) money earns interest at rate r. Assume that

$0 < \beta\delta^{(T+1)-t}(1+r)^{(T+1)-t} < \infty$. *Then, optimal care consumption is*

$$c_{it} = \lambda_{it} + \omega \cdot (1 - P_{it}^{KNPE}) \tag{19}$$

and the relevant price is given by

$$P_{it}^{KNPE} \equiv \beta\delta^{(T+1)-t}(1+r)^{(T+1)-t} \cdot \left(1 - \sum_{\tau=1}^{T} q_{it}(\tau)\right). \tag{20}$$

Here, "KNPE" abbreviates the names of the authors of Keeler et al. (1977) and Ellis (1986). The original result is often cited for showing that the expected end-of-year price is the only relevant price a patient should act on. Proposition 2 shows that this continues to hold under quasi-hyperbolic discounting with interest.

A.5.2 Using the expected end-of-year price as a measure of dynamic incentives

Putting Proposition 1 and 2 side-by-side reveals that in the general case, the relevant price differs across models. It is

$$P_{it} = \begin{cases} 0 & \text{if } P_{it}^{c} = 0 \\ 1 - \beta \cdot \sum_{\tau=t+1}^{T} \delta^{\tau-t} q_{it}(\tau) & \text{if } P_{it}^{c} = 1. \end{cases}$$

in our model and

$$P_{it}^{KNPE} = \beta\delta^{(T+1)-t}(1+r)^{(T+1)-t} \cdot \left(1 - \sum_{\tau=1}^{T} q_{it}(\tau)\right)$$

in the model by Keeler et al. (1977) and Ellis (1986) with quasi-hyperbolic discounting. This is not surprising, as one can in general not summarize complex dynamic incentives using a scalar measure. Also not surprisingly, there is no difference between the two models once patients exhaust the cost sharing limit. Then, the relevant price is zero.

In either of the two models a patient pays an effective price that is given by out-of-pocket costs minus a "bonus". The bonus reflects that spending an additional euro today reduces the remaining deductible for all remaining periods by a euro: as a result of higher spending today, the patient might have to pay less in the future. In fact, the patient will only have to pay less in the future if she crosses the deductible in the future; the probability of such an event occurring at time τ, from the perspective of time t, is given by $q_{it}(\tau)$. Thus, a consumer who spends one additional unit of money out-of-pocket in t will spend one unit of money less in τ with probability $q_{it}(\tau)$.

In our model, spending one euro less in period τ, means not incurring a disutility of 1 euro in period τ with probability $q_{it}(\tau)$. This disutility arises because costs are modeled into flow utility. From the perspective of period t this reduction in disutility in period τ is worth (in expectation) $\beta\delta^{\tau-t}q_{it}(\tau)$. The value of the bonus, in period t, of spending an additional unit is then just the summation of expected disutility reductions over all remaining τs in the year:

$$b_{it} \equiv \beta \cdot \sum_{\tau=t+1}^{T} \delta^{\tau-t} q_{it}(\tau).$$

In the model by Keeler et al. (1977) and Ellis (1986), the inclusion of wealth (with no wealth effects) means that spending money today only affects how much wealth the individual carries over to the next period. The only relevant question, thus, is whether or not the patient will have to pay for the last unit of care within a year. So, the bonus is given by the probability that this is the case, times 1. The value of the bonus, from the perspective of t, is thus

$$b_{it}^{KNPE} \equiv \beta\delta^{(T+1)-t}(1+r)^{(T+1)-t} \cdot \sum_{\tau=t+1}^{T} q_{it}(\tau).$$

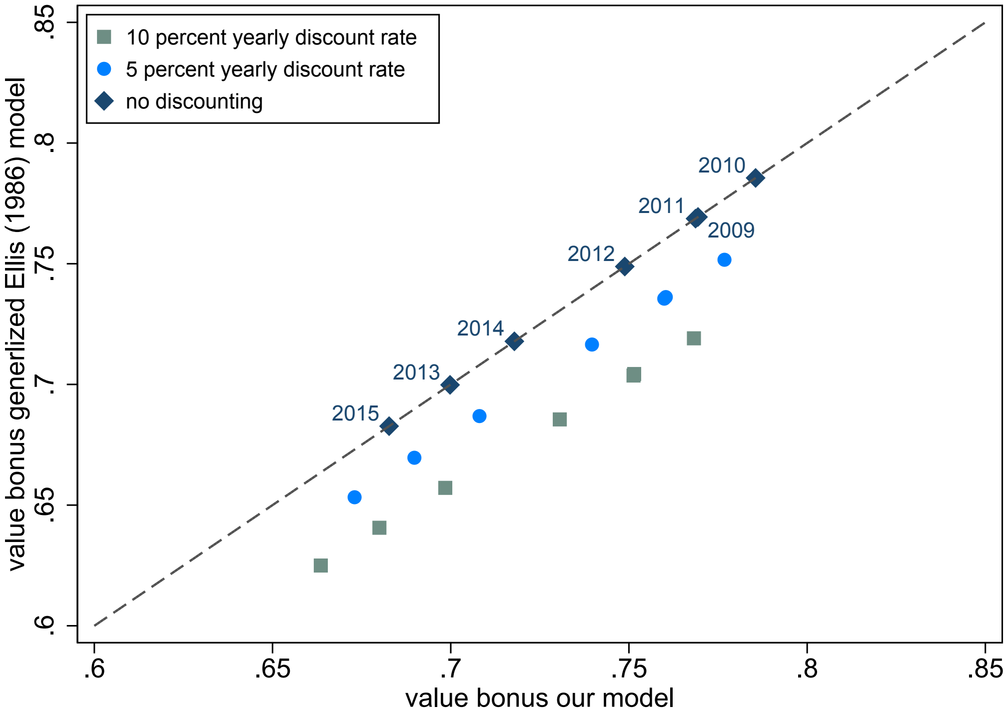
The bonus is the same in both models when $\delta = 1$ and $r = 0$, and the difference between the two is small for the realistic case in which δ is close to 1 and r is small.

When the bonus is the same, then the expected end-of-year price $\bar{P}^e$ is also the same. In our analysis, we relate variation in care consumption to variation in $\bar{P}_{it}^e$ for a given period t. Those changes in $\bar{P}_{it}^e$ are in turn directly related to changes in the true dynamic incentive, b_{it} or b_{it}^{KNPE}. There is no difference between those when $\delta = 1$ and $r = 0$. Also, in both models, according to (10) and (19), the coefficient we would estimate if we would regress healthcare consumption on $\bar{P}_{it}^e$ is $\beta \cdot \omega$, so a reduced-form parameter that measures the reaction to changes in dynamic incentives would have the same meaning in both models.

This remains to hold approximately if $\delta \neq 1$, as long as δ is not too far away from 1 and r is small. To show this, we set $r = 0$ and calculate the average value of the relevant part of the value of the bonus in out model, $\sum_{\tau=t+1}^{T} \delta^{\tau-t} q_{it}(\tau)$, for January of each year (so for $t = 1$) and using the value of $q_{it}(\tau)$ from our data, and plot it against the relevant part of the value of the bonus in the generalized Keeler et al. (1977)-Ellis (1986) model, $\delta^{-1} \cdot \delta^{(T+1)-t} \cdot \sum_{\tau=t+1}^{T} q_{it}(\tau)$. We do so for various values of δ. Here, we add the factor δ^{-1} to make the timing comparable: in our model individuals receive the last at most 11 periods in the future, while in the model with savings they formally receive it 12 periods in the future. The factor δ^{-1} thus makes the two models more comparable. We use $\delta = 1$ for reference, $\delta = 0.996$ that corresponds to a 5 percent yearly discount rate, and $\delta = 0.992$ that corresponds to a 10 percent yearly discount rate.

Figure A.1 shows the result. There are 7 years in our data, which means that for each value of δ we obtain 7 data points. We can see that for each of the 3 values of δ, the values of the bonus from both models are co-monotonic. This means that the ordering is preserved.

Figure A.1: Value of bonus for different models and different values of δ



Note: This figure plots the average value of $\sum_{\tau=2}^{12} \delta^{\tau-1} q_{i1}(\tau)$, which is directly related to the value of the bonus in our model, against $\delta^{12-1} \sum_{\tau=2}^{12} q_{i1}(\tau)$. Done for 3 different discount factors. See text for details.

Moreover, we can see that even for a yearly discount rate of 10 percent, the two values are very close. A regression reveals that a one unit increase in the value of the bonus in our model predicts a 0.894 unit increase in the value of the bonus for the Ellis (1986) model.

A.6 Discussion

In this appendix we have provided a micro foundation for the differences-in-regression-discontinuities analysis we have conducted in the main part of the paper. This shows that our estimate of γ^e has the interpretation of a myopia parameter β times a moral hazard parameter ω. We then show how, based on this, one can perform a counterfactual simulation without solving a structural model.

We have also shown that if we are interested in measuring dynamic incentives, then we will not necessarily have to take a stance on which particular model we prefer. As long as we use a model with quasi-hyperbolic discounting and assume that δ is close to 1 and r is close to zero, our results show that the expected end-of-year "future" price that is also used in the literature based on referencing the particular model by Keeler et al. (1977) is a good measure of dynamic incentives.

Finally, our analysis reveals that a key difference between our model and the original model with wealth by Keeler et al. (1977) is that in our model patients do react to static incentives, with a weight of $(1-\beta)$, while in the model with wealth this is not the case. This difference is not essential for the purpose of measuring the extent to which individuals react to dynamic incentives, but it is a desirable model property as it is more in line with the empirical finding in Keeler and Rolph (1988) and Brot-Goldberg et al. (2017) that patients react to static incentives even conditional on the expected end-of-year price $\bar{P}^e_{it}$. The counterfactual analysis we conduct in Section 8 provides a motivation to prefer the model that we describe in the beginning of this Appendix, as it can be used to study the empirically relevant current price effects within the same model framework. Otherwise, we would predict that patients do not react at all to cost-sharing when they are close to being fully myopic, i.e. for small β that are however strictly positive.[49]

Finally, we show that prior knowledge on the current price effect—which is interesting in many applied contexts, but not the focus of this paper—can be used to separately identify moral hazard effects ω and the myopia parameter β.

[49] In the model with savings, the current price does actually matter when patients are fully myopic, but only then. But this means that the limit of c_{it} as a function of β, for $\beta \to 0$ is not the same as c_{it} when β is exactly zero.

B Inference

B.1 Relating discontinuity sizes to changes in dynamic incentives

Our estimation procedure proceeds in two steps: estimating the discontinuities around the turn of the year and then relating them to the expected end-of-year price. The expected end-of-year price is estimated very precisely, as we estimate it by the fraction of individuals in our sample who hit the deductible. For that reason, we ignore the associated estimation error and focus on accounting for the estimation error in our estimates of the discontinuities when calculating the standard errors in the second step.

Following Hanushek (1973), the regression equation for the second step is given by

$$\hat{\Delta}_y = \alpha + \beta \bar{P}^e_{1,y} + (\nu_y + u_y) \tag{21}$$

where $\hat{\Delta}_y$ is the estimated change in care consumption for year-pair $\{y-1, y\}$, u_y is the estimation error that arises from using the estimates $\hat{\Delta}_y$ as the dependent variable and ν_y is the error term for the regression equation that uses the actual unobserved discontinuities Δ_y as the dependent variable. We assume ν_y and u_y are independent, ν_y is homoskedastic with $Var(\nu_y) = \sigma^2$, $Var(u_y) = e_y^2$ and $Cov\left(u_y, u_{y'}\right) = 0 \ \forall \ y \neq y'$. $\sqrt{e_y^2}$ corresponds to the standard error of the estimated discontinuity $\hat{\Delta}_y$.

Let $\varepsilon_y = \nu_y + u_y$. Since the ε_y is a sum of 2 independent random variables, and $Cov(u_y, u_{y'}) = 0 \ \forall \ y \neq y'$, we have that the variance-covariance matrix of ε_y is given by

$$\mathbb{E}\left[\varepsilon\varepsilon'\right] = \begin{bmatrix} \sigma^2 + e_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2 + e_Y^2 \end{bmatrix}.$$

Since u_y is not restricted to be the same across $\hat{\Delta}_y$, having such a model results in heteroskedastic errors. Ignoring such heteroskedasticity would thus lead to inefficient OLS estimates and incorrect conclusions from statistical tests. A simple way to deal with this heteroskedasticity is use a generalized least squares (GLS) estimator. Since σ^2 is not known, however, we have to use feasible GLS (FGLS) instead. For this, we replace σ^2 by an estimate.

To estimate σ^2, we first run an OLS regression of $\hat{\Delta}_y$ on $\bar{P}^e_{1,y}$, essentially ignoring the finite-sample error from our estimates $\hat{\Delta}_y$. Hanushek (1973, p. 67) shows that the residuals from this regression has an expected variance that is a function of σ^2 and e_y^2, which then allows him to solve for σ^2. This yields an estimate of σ^2. This estimate, $\hat{\sigma}^2$, is then used, in a next step, to run a weighted least-squares regression with weights, w_y, given by

$$w_y = \frac{1}{\hat{\sigma}^2 + e_y^2}.$$

B.2 Non-parametric monotonicity tests

We use the MR and Up/Down tests that have been proposed in Patton and Timmermann (2010) as non-parametric tests for monotonicity. Patton and Timmermann (2010) argue that these test statistics can provide strong evidence for an economic theory. It was also the only non-parametric test of monotonicity we could find that made use of the estimated coefficients rather than just their ranking.

Consider our 7 estimates of changes in expenditures around the turn of the year, $\hat{\Delta}_y$, ordered with respect to the expected end-of-year price, with the smallest expected end-of-year price first and denote them by $\tilde{\Delta}_1, \ldots, \tilde{\Delta}_7$. Economic theory posits that years with a higher expected end-of-year price should exhibit a larger decrease in percentage spending. If there was no estimation error, this would mean

$$\tilde{\Delta}_1 > \tilde{\Delta}_2 > \ldots > \tilde{\Delta}_7. \tag{22}$$

Let d_i denote the difference between consecutive ranks i and $i+1$ in (22). Consider the minimum of all d_i's. If this minimum is larger than 0 (in a statistical sense), it must be the case that every other difference is larger than 0. This provides the basis for the MR test statistic

$$J_T = \min_i \{d_i\}. \tag{23}$$

Since the asymptotic distribution of J_T does not have a closed form solution, we have to obtain p-values via bootstrapping. This is done, for each bootstrap repetition, by drawing from the distribution of the estimated coefficients, ordering them with respect to the expected end-of-year price, and then computing J_T^b, which is the value of the MR test statistic for bootstrap repetition b. To impose the null hypothesis, we then subtract the quantity in (23), J_T, from J_T^b. The p-value is given by

$$\frac{1}{B}\sum_{b=1}^{B} 1\left\{\left(J_T^b - J_T\right) > J_T\right\}.$$

As mentioned in Section 5.2 and discussed also in Patton and Timmermann (2010), the MR test can lack power. For our specific application, this lack of power stems from 2 reasons:

1. There is a cardinality in the ranking that is not taken into account. For example, the change in expected end-of-year price from 2012 to 2013 is rather large, but the test statistic treats all of these changes the same.

2. The magnitude of the change in coefficients is not taken into account. For example, the estimated coefficients changes by quite a large amount from 2012 to 2013.

To diagnose whether the test statistic fails to reject the null of no monotonic relationship due to a lack of power, Patton and Timmermann (2010) propose the Up and Down test statistic,

$$J_{Up} = \sum_{i=1}^{N} |d_i| \{d_i > 0\} \tag{24}$$

$$J_{Down} = \sum_{i=1}^{N} |d_i| \{d_i < 0\}. \tag{25}$$

(23) picks up significant changes that are in line with our expectations. (25) picks up significant changes that go against expectations. Therefore, we expect the p-value for the Up test to be small and the p-value for the Down test to be above conventional levels of significance. We obtain p-values for both (23) and (25) using the same bootstrap procedure as for the MR test.

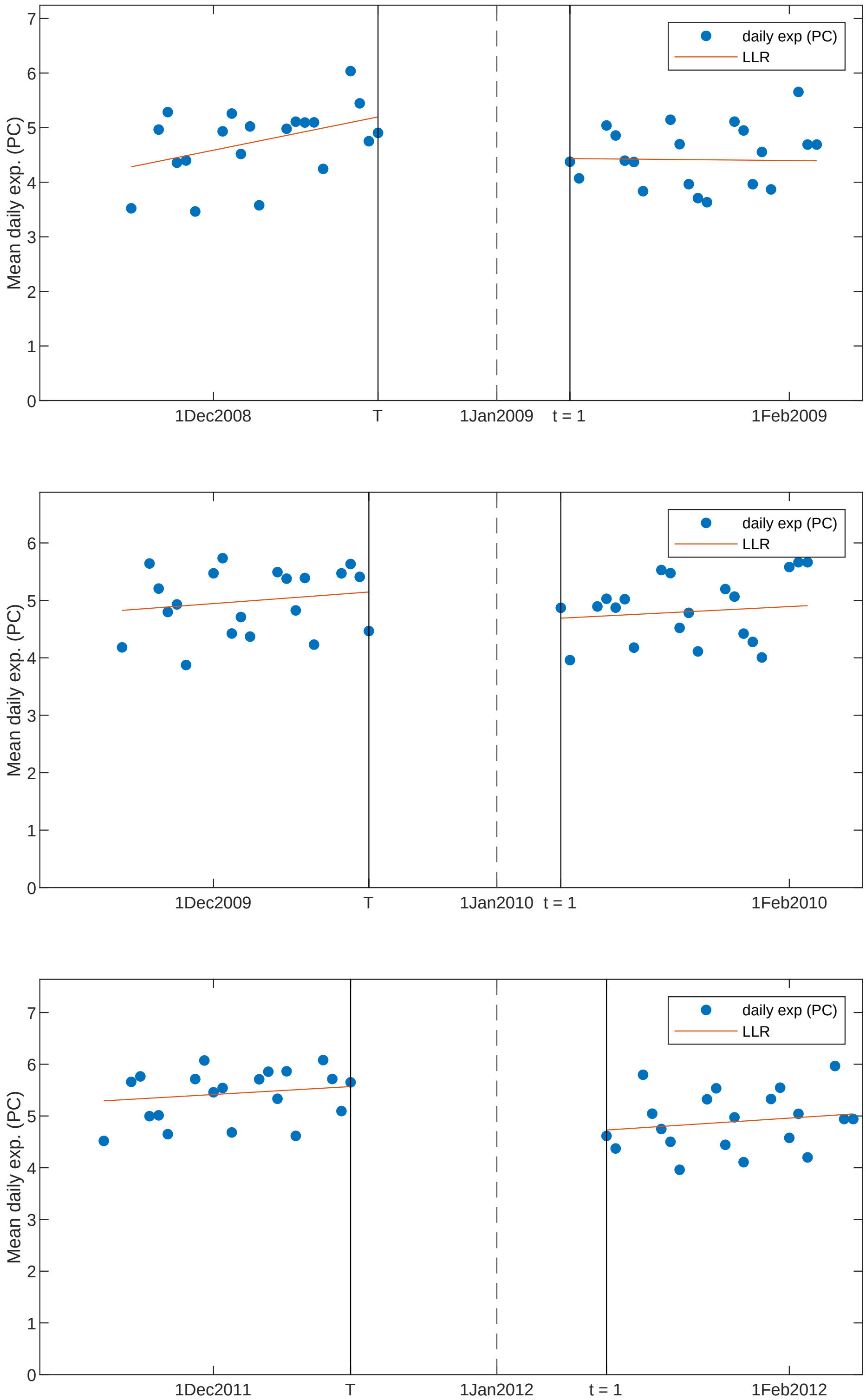
C Additional tables and figures that we refer to in the text

Table A.1: Dates used in analysis

year-pair	last day in year y	first day in year $y+1$
2008 – 2009	18^{th} December 2008	8^{th} January 2009
2009 – 2010	17^{th} December 2009	7^{th} January 2010
2010 – 2011	16^{th} December 2010	6^{th} January 2011
2011 – 2012	15^{th} December 2011	12^{th} January 2012
2012 – 2013	20^{th} December 2012	10^{th} January 2013
2013 – 2014	19^{th} December 2013	9^{th} January 2014
2014 – 2015	18^{th} December 2014	8^{th} January 2015

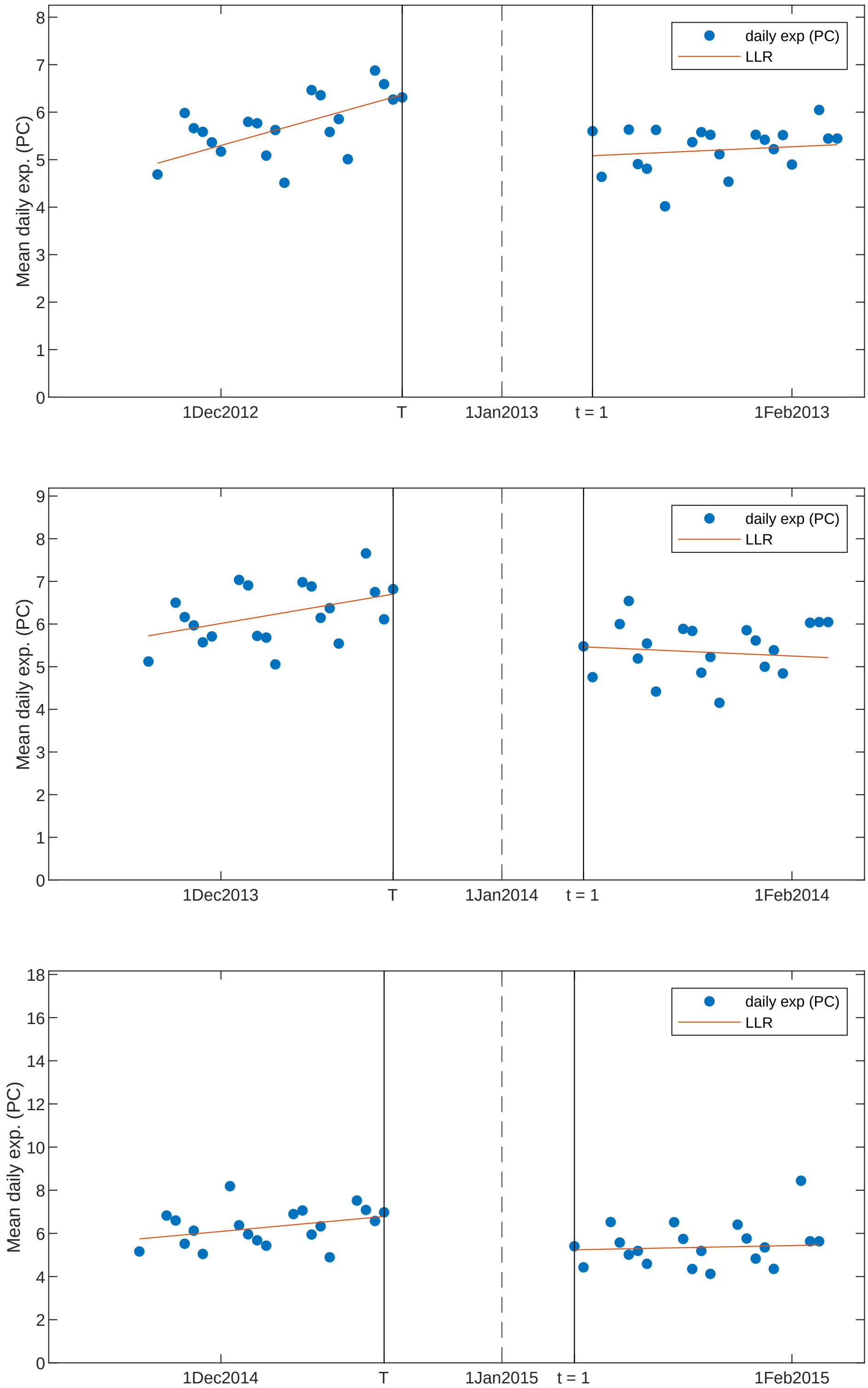
Note: In our analysis we use the last 20 regular days of year y and the first 20 regular days of year $y+1$. Regular days exclude the weekend. We omit a set of days around the turn of the year that include the Christmas break and the actual turn of the year. This table shows the last day in year y that we use for the analysis, and the first day in year $y+1$. ToY refers to turn of year. The relevant holidays before the turn of the year is the 24^{th} of December, while the relevant holiday after the turn of the year is the 1^{st} of January.

Figure A.2: RD figures for year-pairs 2008/2009, 2009/2010, and 2011/2012



Notes: This figure replicates Figure 3, which is for the year-pair 2010/2011, for other year-pairs.

Figure A.3: RD figures for year-pairs 2012/2013, 2013/2014, and 2014/2015



Notes: This figure replicates Figure 3, which is for the year-pair 2010/2011, for other year-pairs.

Table A.2: Unemployment rate and change in purchasing power over time

Year	unemployment rate	change in purchasing power
2008	3.7%	1.4%
2009	4.4%	1.9%
2010	5.0%	-0.5%
2011	5.0%	-0.7%
2012	5.8%	-1.1%
2013	7.3%	-1.2%
2014	7.4%	1.8%
2015	6.9%	1.1%

Notes: Taken from the Statistics Netherlands website. See https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80590ned/table?ts=162 for the unemployment rate and https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70959ned/table?ts=162 for the change in purchasing power (both accessed August 2021). The unemployment rate is for individuals of age 15 to 75, yearly, and not seasonally adjusted. The change in purchasing power is based on total income.

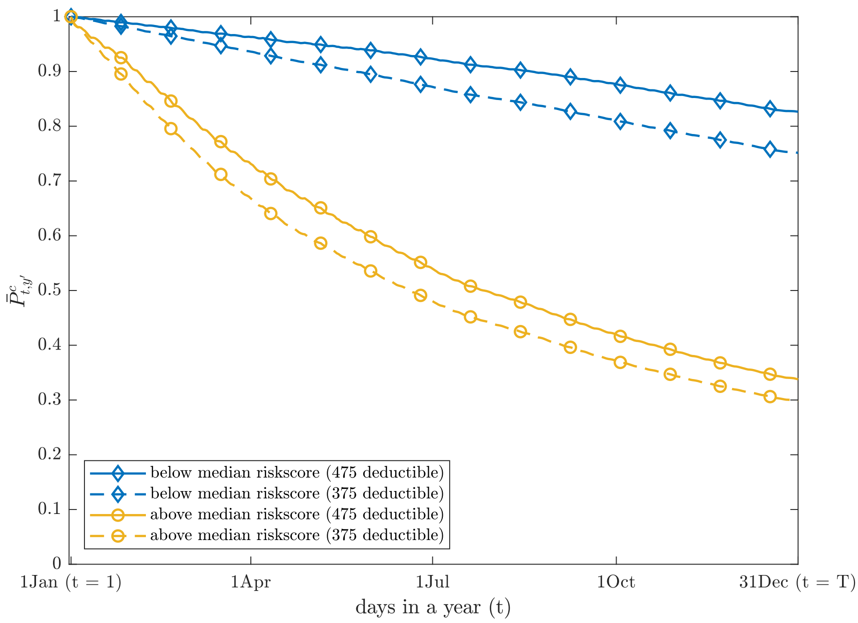
Table A.3: Expenditure deflator values

Year	own deflator	Eurostat
2008	.665	0.782
2009	.719	0.785
2010	.739	0.800
2011	.776	0.842
2012	.831	0.933
2013	.908	0.991
2014	.925	1.00
2015	1	1

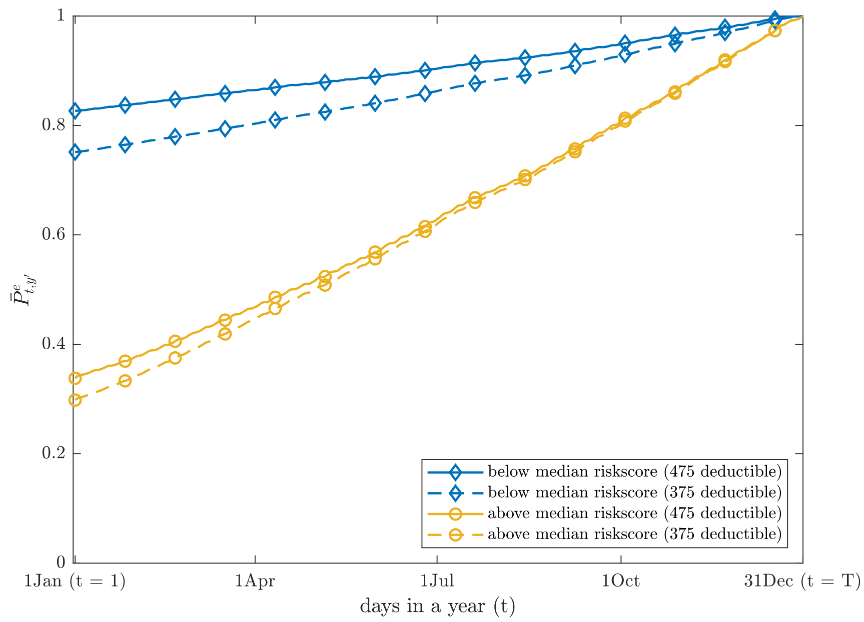
Notes: This table shows two alternative deflators for health care spending. Deflator values for the own deflator were obtained from mean expenditures (only weekdays) from the 37^{th} week of the year to the end of the year, when individuals in our sample face no price of care. See Section 7.1 for details.

Figure A.4: Effects of increasing the deductible by €100 on static and dynamic incentives

(a) Fraction individuals with current price of 1



(b) Expected end-of-year price for those who still have a current price of 1



Notes: Figure (a) shows the share of individuals with a positive current price at time t, $\bar{P}^c_{t,y}$, for two deductible levels, €375 and €475, across riskscore. Figure (b) depicts the predicted average expected end-of-year price for individuals with a positive current price at time t, $\bar{P}^e_{t,y}$, for two deductible levels, €375 and €475, across riskscore. See Section 8 for details.

D Additional robustness checks

This appendix contains a number of additional robustness checks. Table A.4 summarizes the results. Some of these are already reported in Table 6. Here, we add the extensive margin. Corresponding figures are reported separately in Appendix E.

Row 1 of Table A.4 shows the result from a log specification. We added 1 to daily expenditures before taking the log.

Row 2 shows the results when we pseudo-censor health care expenditures at 5000 instead of 500.

In Section 7.1, in line with the predictions of our model, we saw that patients do not respond to our measure of dynamic incentives for care that costs more than €500. As an additional robustness check, in row 3, we also report the results for care that costs of at least €5000.

Row 4 and 5 show results for big spending events.

Row 6 and 7 show results for individuals of age 67 and above.

In rows 8 and 9, we use raw averages before and after the change of the year. We include weekends in row 10 and 11.

We also run our empirical approach on weekly level data. We create this weekly level data set through the following steps:

1. We create the daily level final data set, with all the relevant dates for the donut hole.
2. We start counting 7 days from the day before the turn of the year (our T in the RD). This is one week for days before the turn of the year.
3. We start counting 7 days from the day after the turn of the year (our $T+1$ in the RD). This is one week for days after the turn of the year.
4. We sum expenditures on the weekend within these 7 days and pseudo-censor it at €500.

Then, we run our analysis on this dependent variable. Results are reported in row 12 and 13.

Rows 14 to 19 show results for alternative donut holes.

Rows 20 and 21 show results for a placebo test. In the Netherlands only individuals above the age of 18 (inclusive) are subject to cost-sharing in the form of deductibles, while the price of care for individuals below the age of 18 is always zero. Individuals below the age of 18, thus, provide a simple placebo test for our empirical methodology—since they do not face any price of care for all the years in consideration, their changes in healthcare utilization should not exhibit an increasing relationship with $\bar{P}^e_{1,y+1}$.[50] It is important to mention that dental care was removed from estimating the changes from resetting contracts since it is not covered in the basic package for our main sample.

[50] Since these individuals do not face the deductible, we compute their $\bar{P}^e_{1,y+1}$ as the 1 minus the proportion of total individuals that would have crossed the actual deductible in year $y+1$.

Table A.4: Additional robustness checks

specification		γ^e		MR test	Up test	Down test
1	log dependent variable	-0.414	(0.0738)	0.649	0.000	0.694
2	PC at 5000	-13.119	(4.7653)	0.028	0.267	0.959
3	spending greater than €5000	0.000	(0.0009)	0.586	0.757	0.203
4	sample of people with large spending in y	-12.884	(1.8245)	0.151	0.057	0.953
5	sample of people with large spending in y (ext)	-0.092	(0.0213)	0.062	0.044	0.987
6	greater than 67 years of age	-33.758	(5.9674)	0.538	0.023	0.600
7	greater than 67 years of age (ext)	-0.232	(0.0609)	0.069	0.004	0.936
8	uniform kernel, w/o linear RD functional form	-13.009	(2.9124)	0.002	0.001	0.999
9	uniform kernel, w/o linear RD functional form (ext)	-0.070	(0.0328)	0.619	0.000	0.262
10	including weekends	-11.054	(0.9368)	0.000	0.021	1.000
11	including weekends (ext)	-0.100	(0.0136)	0.392	0.000	0.882
12	weekly data, only weekends	-10.167	(2.6211)	0.035	0.079	0.973
13	weekly data, only weekends (ext)	-0.031	(0.0185)	0.998	0.000	0.036
14	Thursday, 27 day donut	-12.576	(2.8153)	0.477	0.000	0.598
15	Thursday, 27 day donut (ext)	-0.126	(0.0408)	1.000	0.000	0.001
16	Wednesday, 20 day donut	-8.350	(1.3796)	0.000	0.147	1.000
17	Wednesday, 20 day donut (ext)	-0.042	(0.0322)	0.991	0.003	0.078
18	Wednesday, 27 day donut	-10.742	(2.2300)	0.000	0.051	1.000
19	Wednesday, 27 day donut (ext)	-0.105	(0.0519)	1.000	0.000	0.005
20	placebo (under 18 years of age)	1.215	(2.3187)	0.225	0.729	0.508
21	placebo (under 18 years of age) (ext)	-0.002	(0.0105)	0.172	0.747	0.632
22	last year-pair removed	-9.634	(1.5556)	0.000	0.009	0.998
23	last year-pair removed (ext)	-0.109	(0.0255)	0.411	0.000	0.832

Notes: The table replicates Table 3 for different specifications. See Appendix D for details. In the tables, (ext) stands for results at the extensive margin.

Finally, in footnote 13, we mentioned that the data were collected for a project in which a new payment model for GPs was evaluated. This pilot began in July 2014 and could thus affect our estimated discontinuity size for the year-pair $\{2014, 2015\}$. To ensure that this one year-pair is not the sole driver of our findings, we apply our empirical approach on all year-pairs other than the year-pair $\{2014, 2015\}$. The results are shown in row 22 and 23.

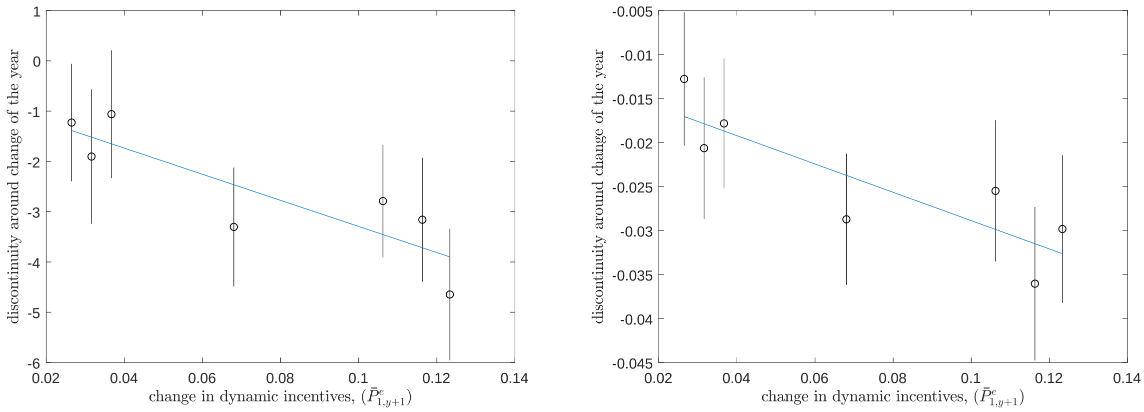
E Additional figures on the dependence of discontinuity sizes on dynamic incentives

This appendix replicates Figure 4 for alternative outcomes and samples.

Figure A.5: Results for males

(a) PC expenditures

(b) extensive margin

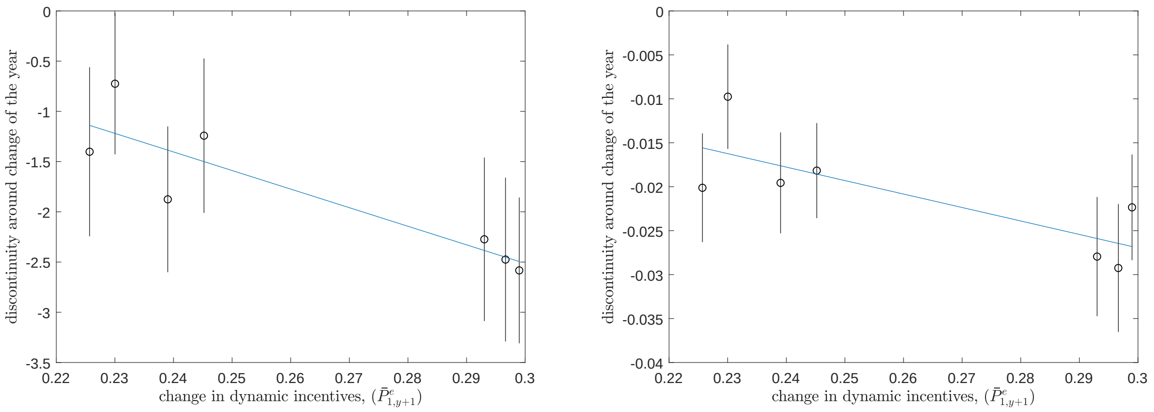


Notes: This figure replicates Figure 4 for males.

Figure A.6: Results for females

(a) PC expenditures

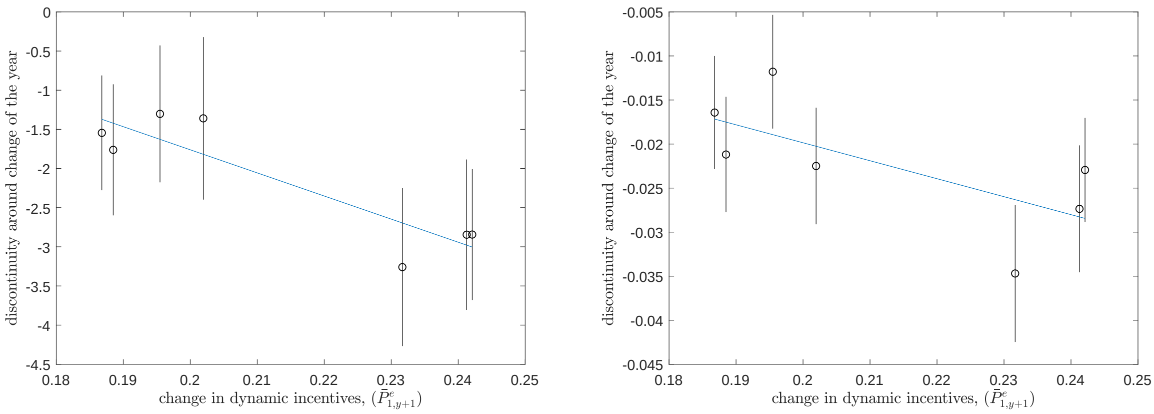
(b) extensive margin



Notes: This figure replicates Figure 4 for females.

Figure A.7: Results for below median income

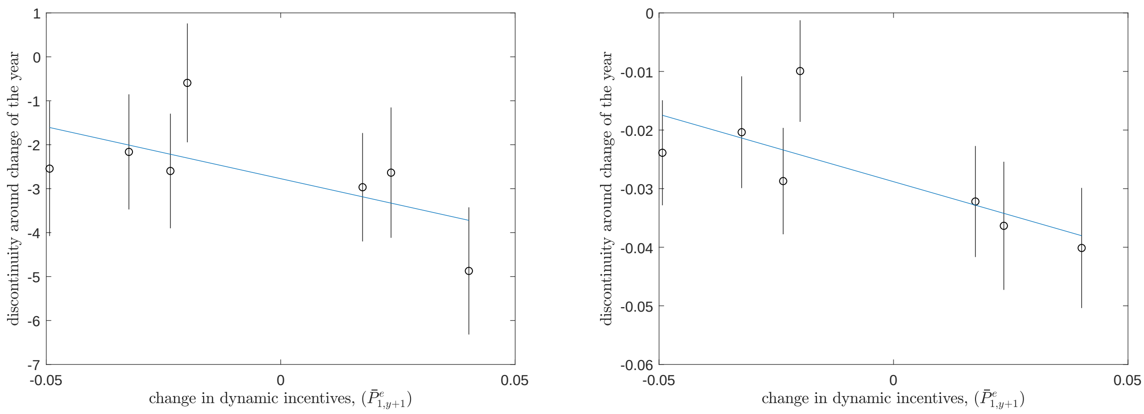
(a) PC expenditures (b) extensive margin



Notes: This figure replicates Figure 4 for individuals who live in below median income neighborhoods.

Figure A.8: Results for above median income

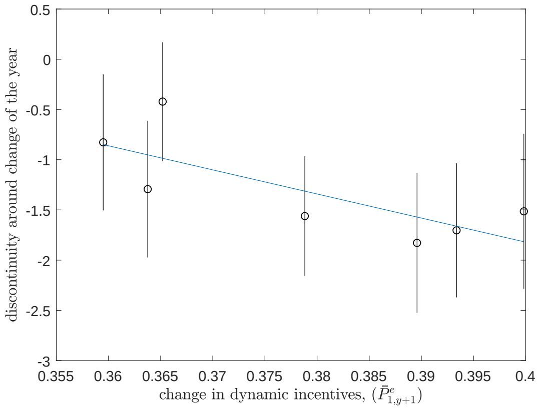
(a) PC expenditures (b) extensive margin



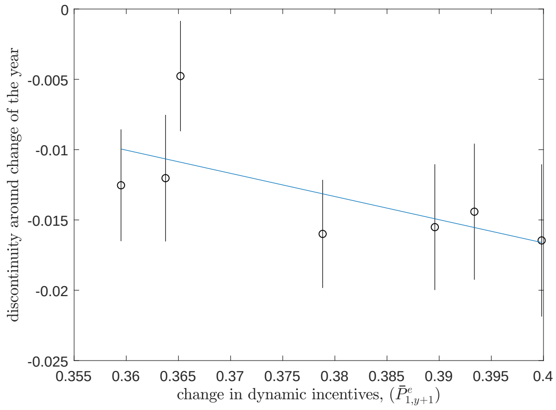
Notes: This figure replicates Figure 4 for individuals who live in above median income neighborhoods.

Figure A.9: Results for age below 45 years

(a) PC expenditures



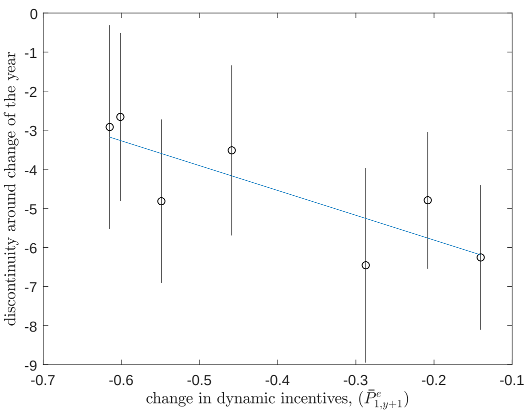
(b) extensive margin



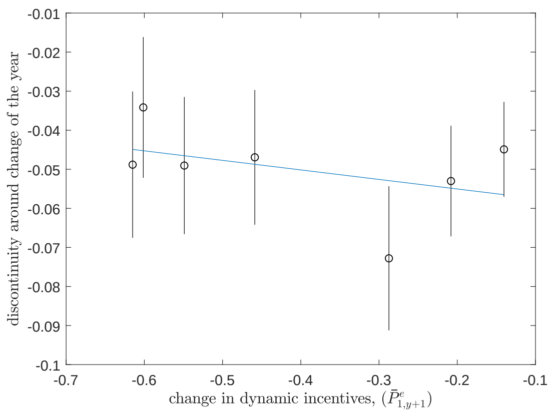
Notes: This figure replicates Figure 4for individuals below 45 years of age.

Figure A.10: Results for age above 45 years

(a) PC expenditures



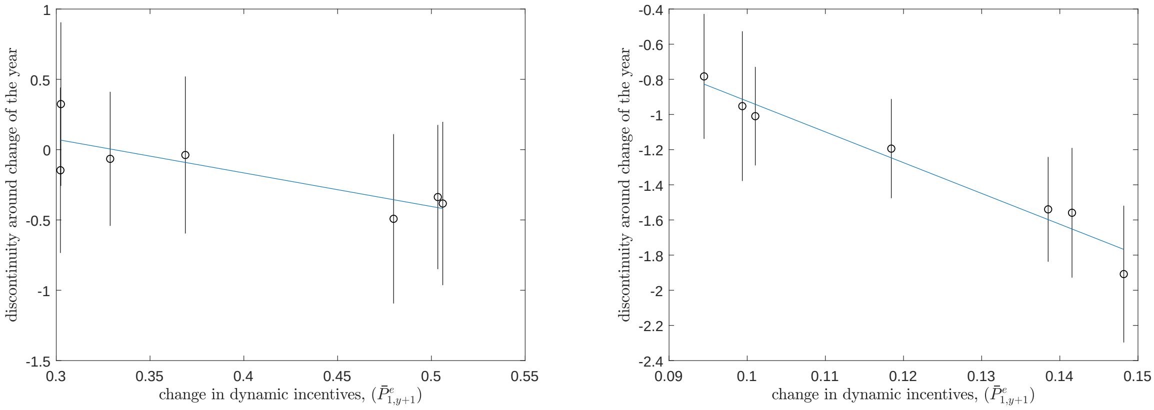
(b) extensive margin



Notes: This figure replicates Figure 4 for individuals above 45 years of age.

Figure A.11: Results by riskscore group (PC expenditures)

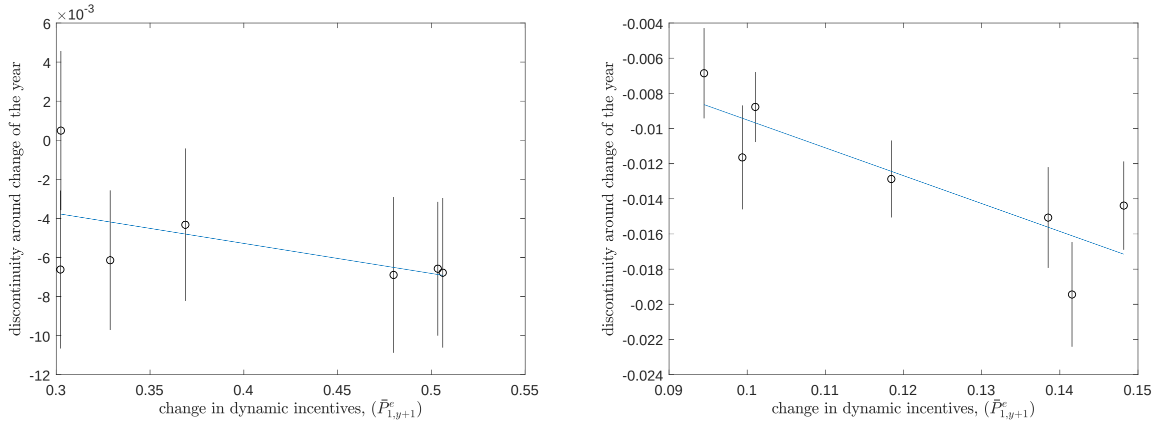
(a) Below median riskscore (b) Above median riskscore



Notes: This figure replicates Figure 4(a) by riskscore group.

Figure A.12: Results by riskscore group (extensive margin)

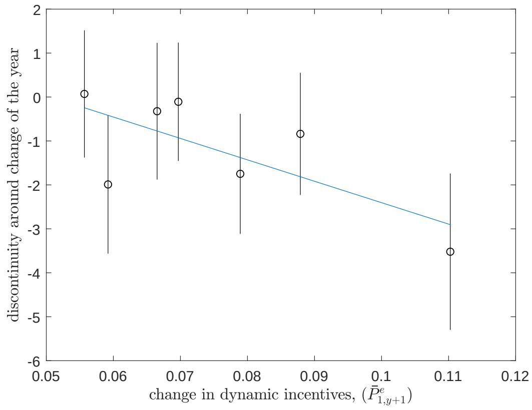
(a) Below median riskscore (b) Above median riskscore



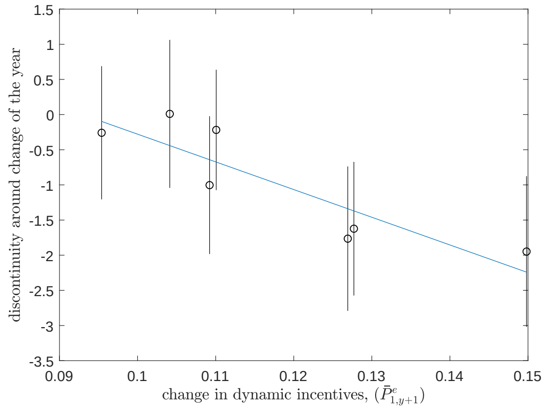
Notes: This figure replicates Figure 4(b) by riskscore group.

Figure A.13: Results by position in spending distribution (PC expenditures)

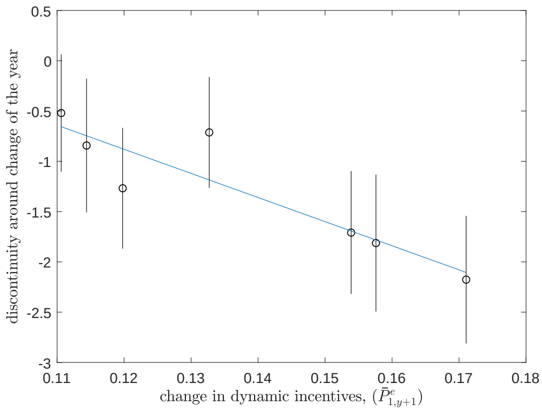
(a) Top 5% of spenders



(b) Top 10% of spenders



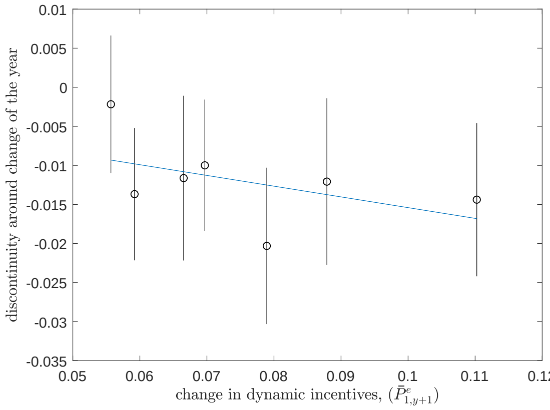
(c) Top 20% of spenders



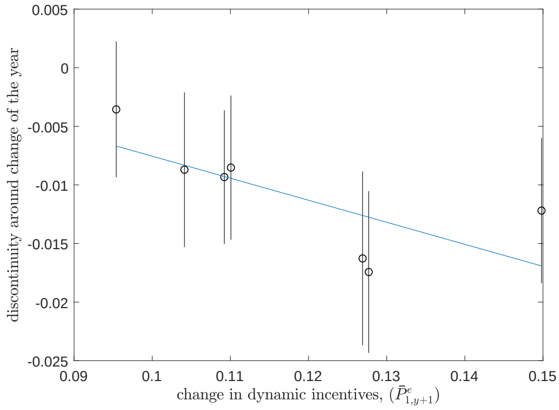
Notes: This figure replicates Figure 4(a) for the top 5%, 10%, and 20% spenders.

Figure A.14: Results by position in spending distribution (extensive margin)

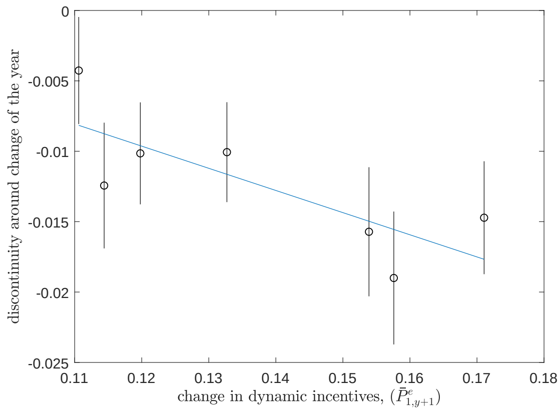
(a) Top 5% of spenders



(b) Top 10% of spenders

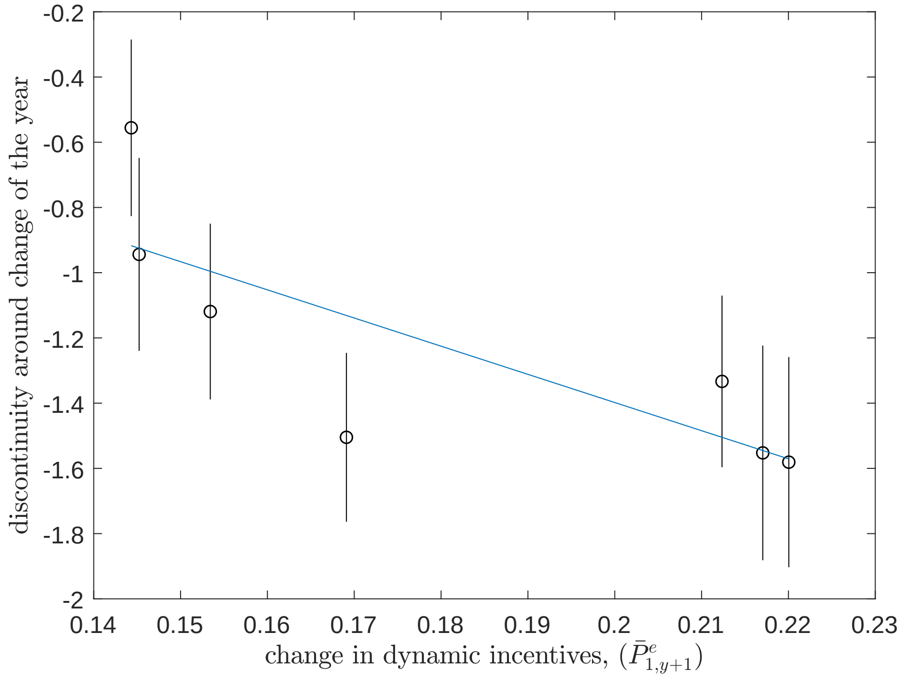


(c) Top 20% of spenders



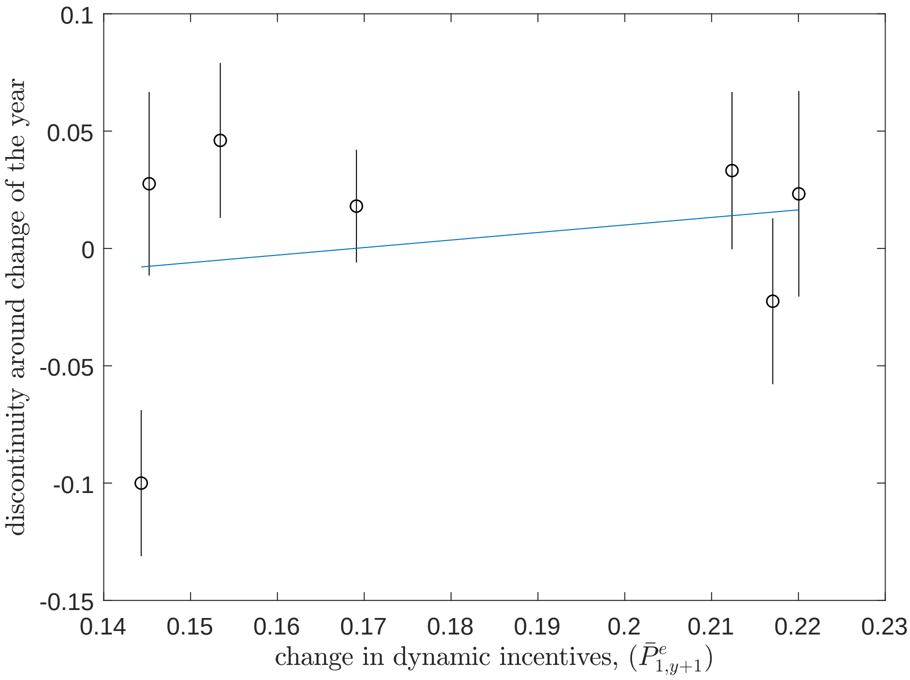
Notes: This figure replicates Figure 4(b) for the top 5%, 10%, and 20% spenders.

Figure A.15: Eurostat-deflated expenditures



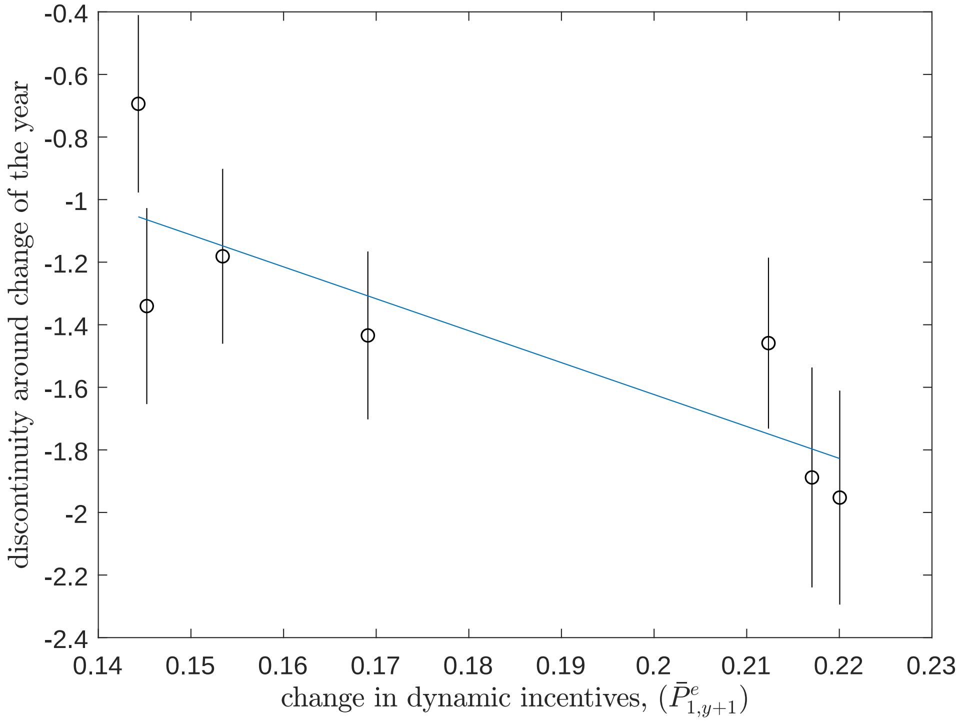
Notes: This figure replicates Figure 4(a) when we deflate expenditures using the Eurostat deflator.

Figure A.16: Spending on GP care



Notes: This figure replicates Figure 4(a) for GP care.

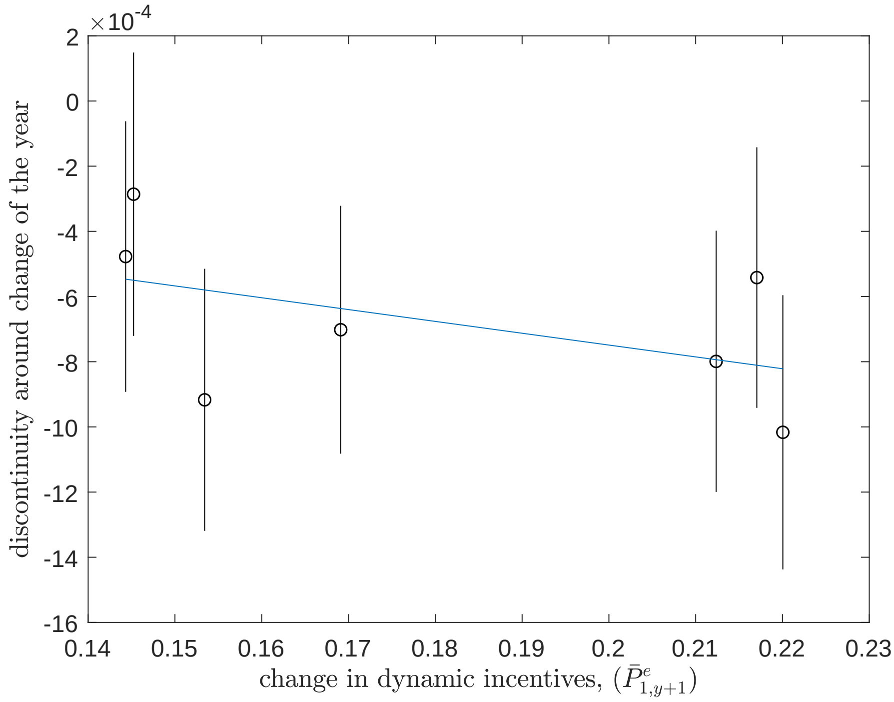
Figure A.17: Deflated expenditures using own deflator



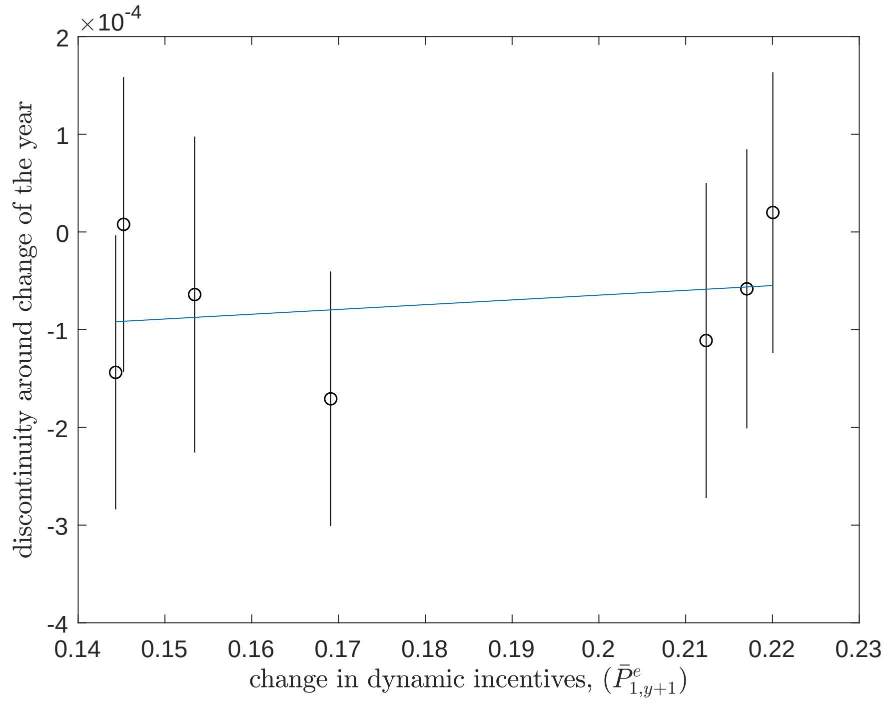
Notes: This figure replicates Figure 4(a) when we deflate expenditures using the our own deflator.

Figure A.18: Relationship between jump and $\bar{P}^e_{1,y+1}$ (large spending events)

(a) Prob. spending > €500



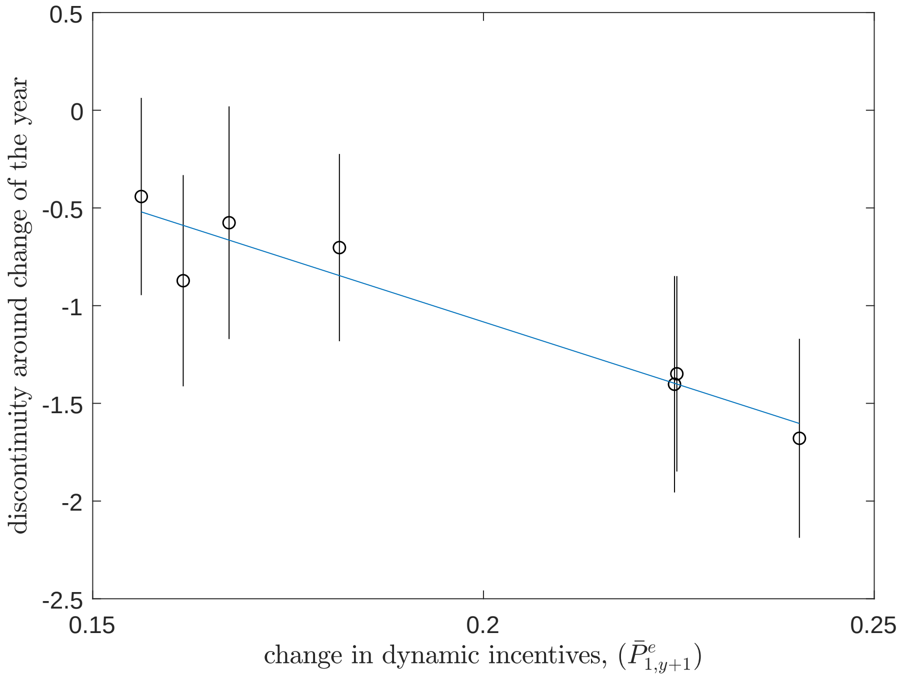
(b) Prob. spending > €5000



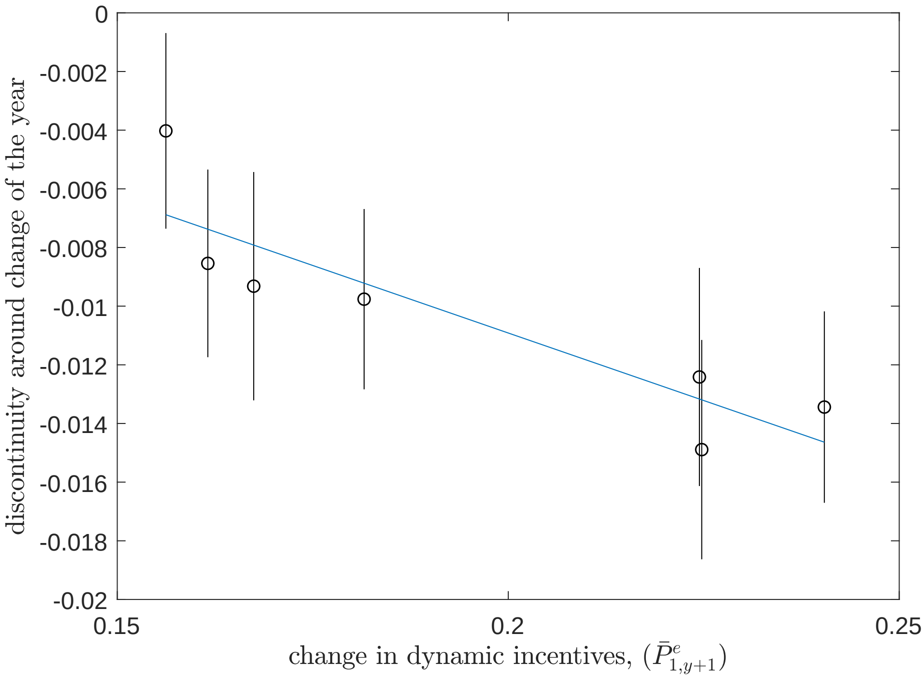
Notes: This figure replicates Figure 4 for spending events larger than €500 and €5000 respectively.

Figure A.19: Relationship between jump and $\bar{P}^e_{1,y+1}$ (alternative sample restriction)

(a) Pseudo-censored expenditures



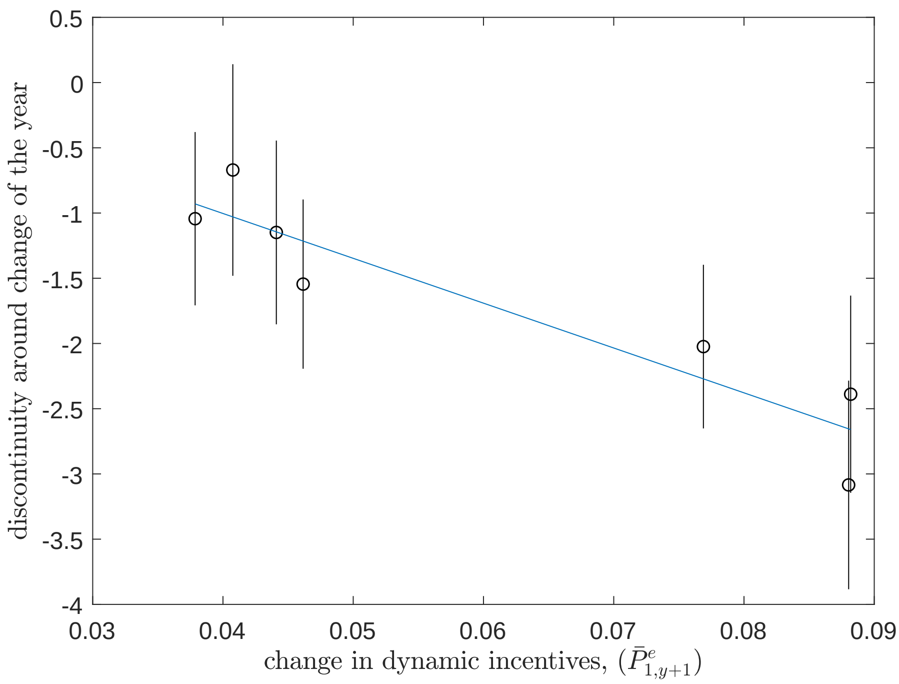
(b) Prob. any exp.



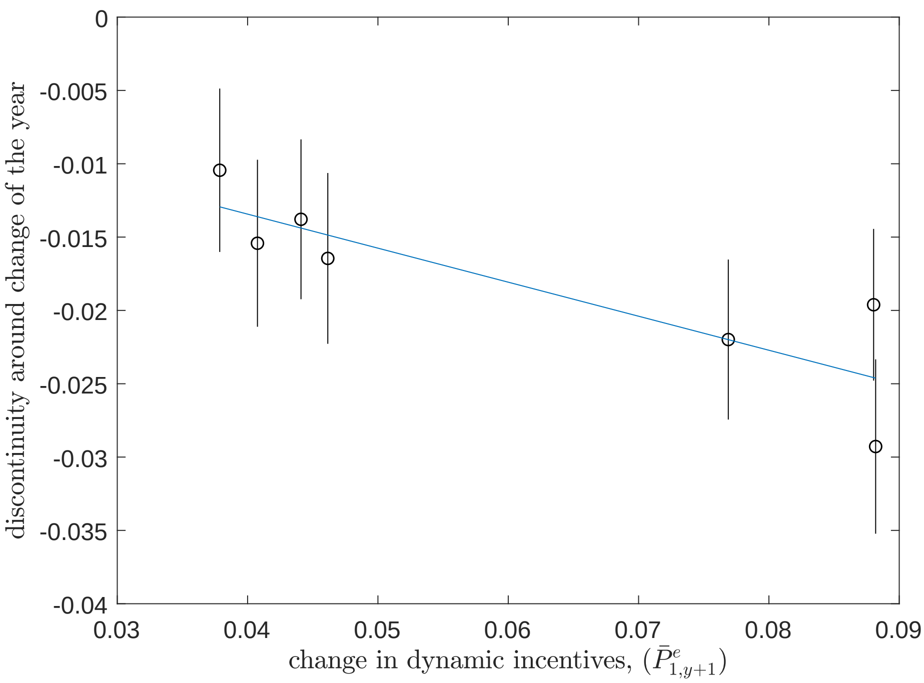
Notes: This Figure replicates Figure 4 when we add the restriction that individuals in our sample must have at least one spending event greater than €500 before August of year *y*.

Figure A.20: Relationship between jump and $\bar{P}^e_{1,y+1}$ (>= 67 years old)

(a) Pseudo-censored expenditures



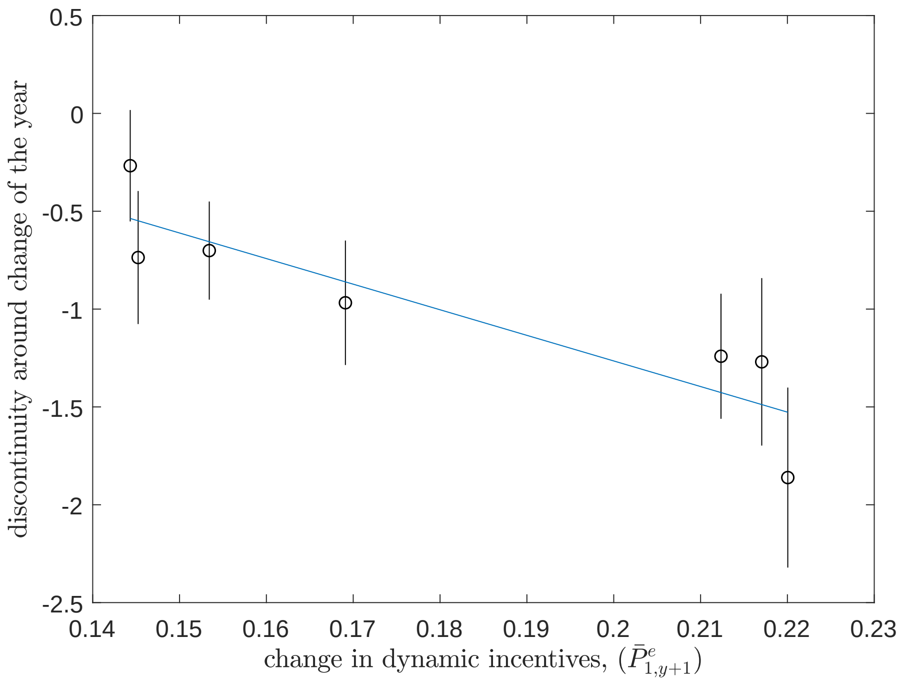
(b) Prob. any exp.



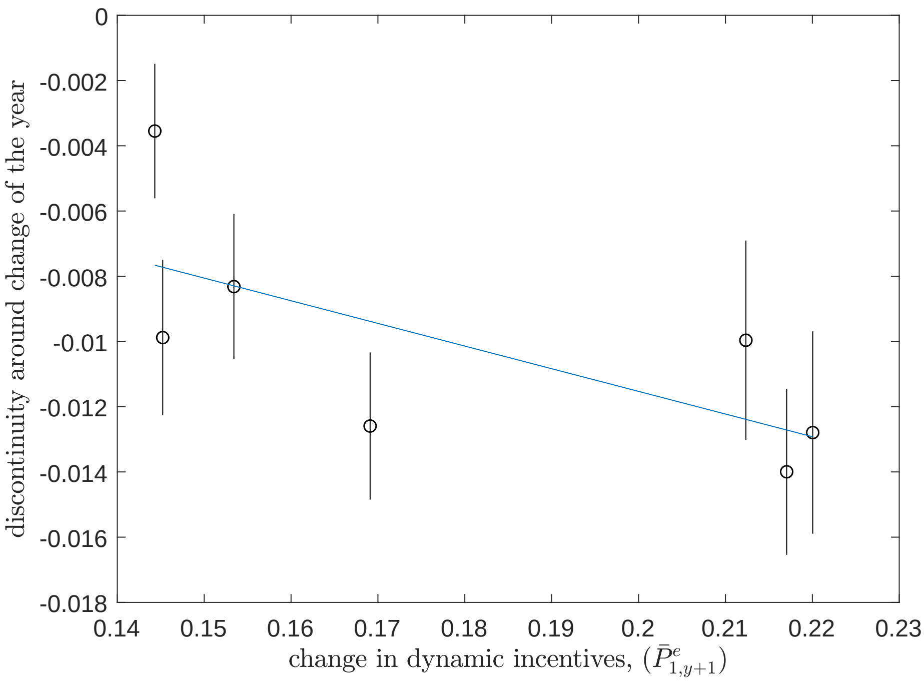
Notes: This figure replicates Figure 4 for individuals of age 67 or older.

Figure A.21: Relationship between jump and $\bar{P}^{e}_{1,y+1}$ (uniform kernel, w/o linear RD functional form)

(a) Pseudo-censored expenditures



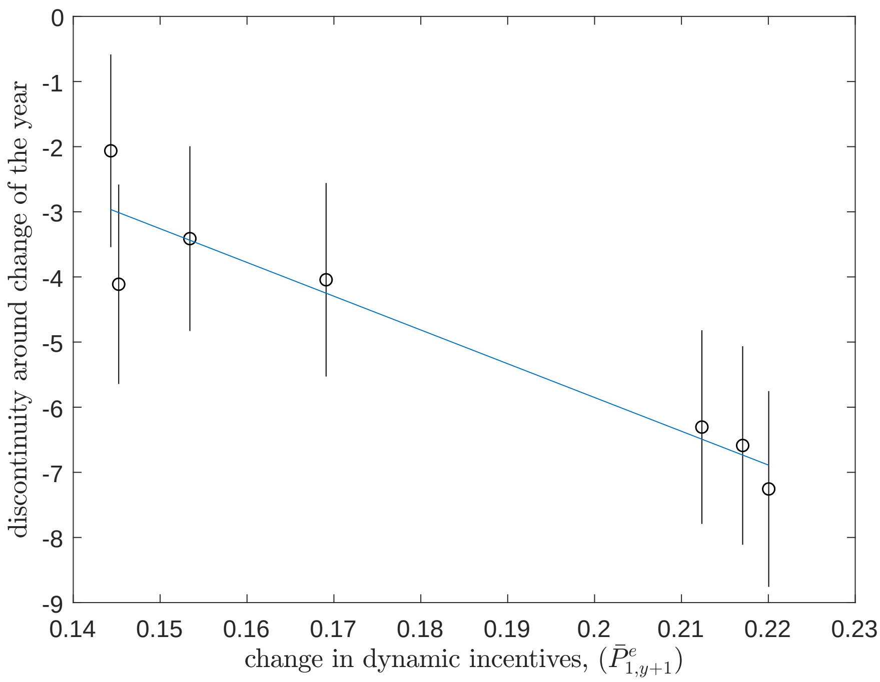
(b) Prob. any exp.



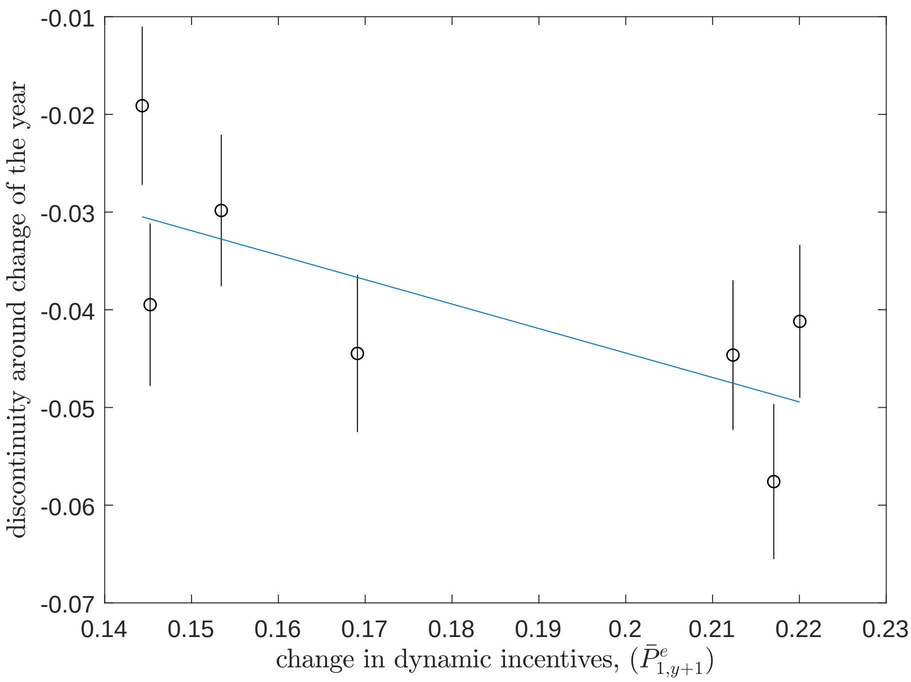
Notes: This figure replicates Figure 4 when we use a uniform kernel and a polynomial of order 0.

Figure A.22: Relationship between discontinuity sizes and $\bar{P}^e_{1,y+1}$ (weekly data)

(a) Pseudo-censored expenditures



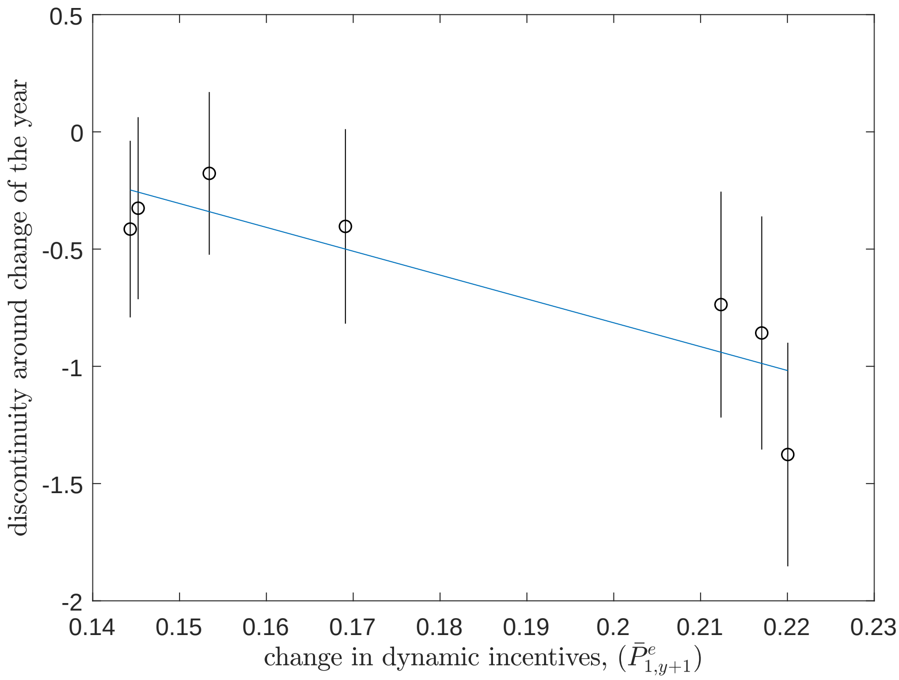
(b) Prob. of weekly exp.



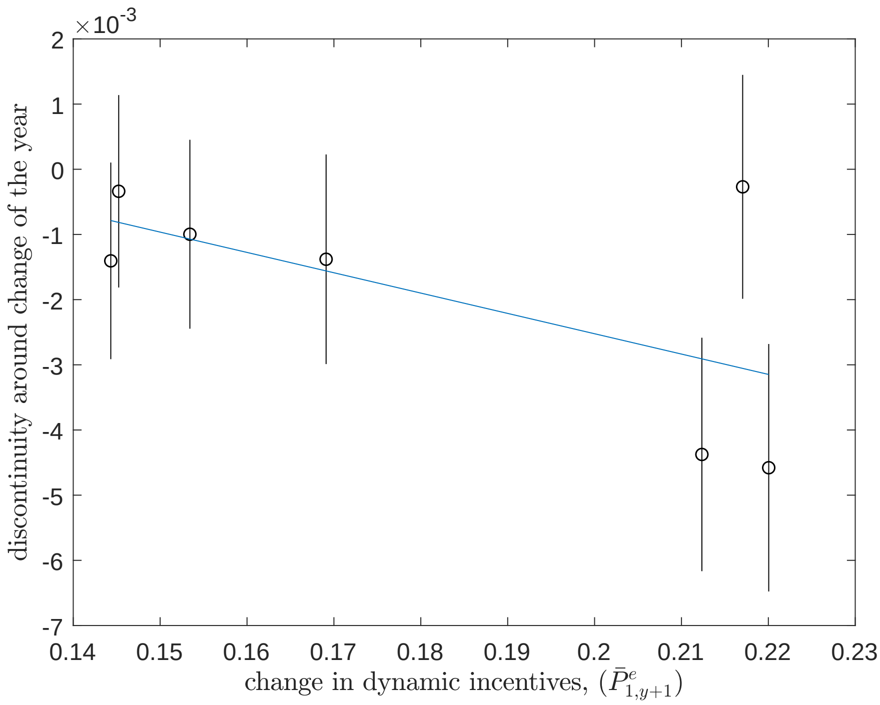
Notes: This table replicates Figure 4 when we use weekly data.

Figure A.23: Relationship between discontinuity sizes and $\bar{P}^e_{1,y+1}$ (weekly data, only weekend care included)

(a) Pseudo-censored expenditures



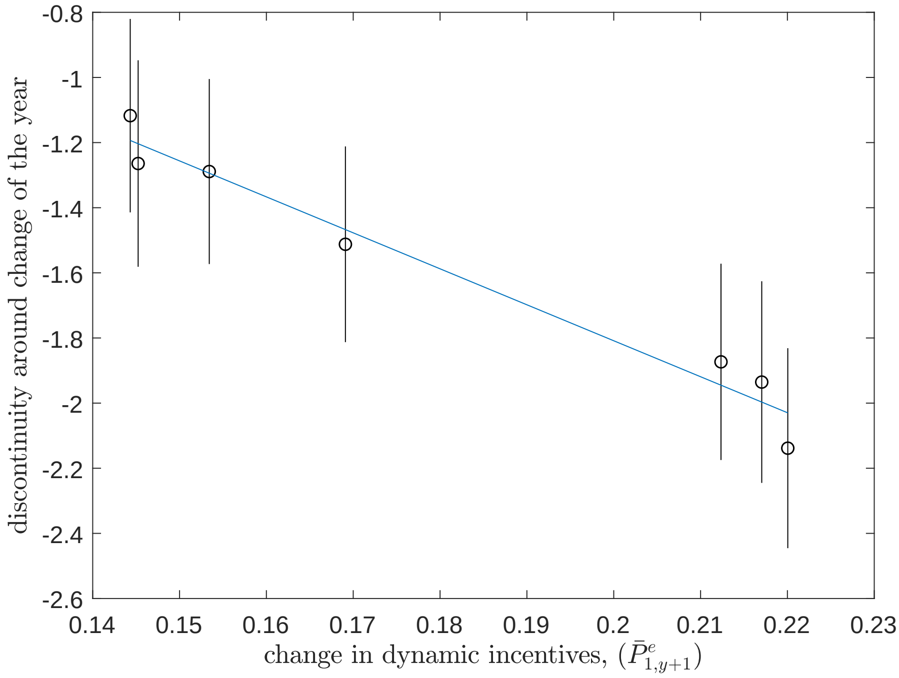
(b) Prob. of weekly exp.



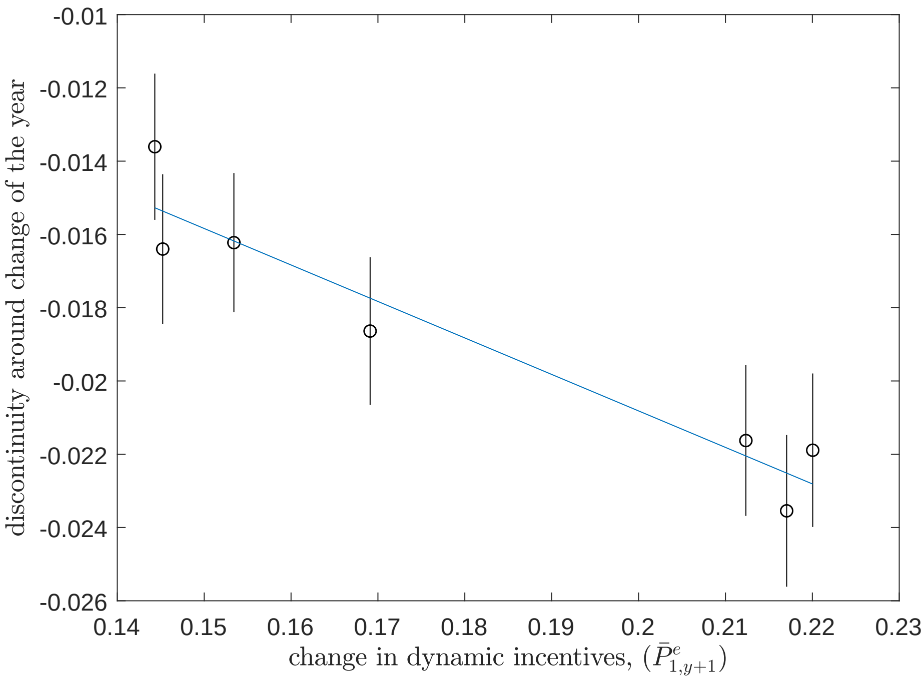
Notes: This table replicates Figure 4 when we use weekly data that sums over only weekend care.

Figure A.24: Relationship between jump and $\bar{P}^{e}_{1,y+1}$ (weekends included)

(a) Pseudo-censored expenditures

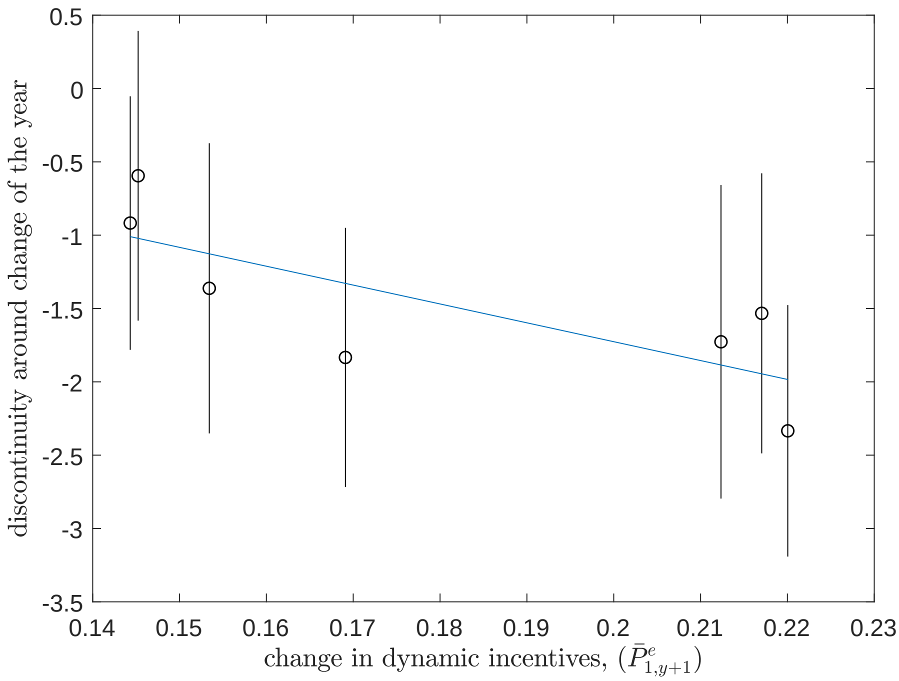


(b) Prob. any exp.



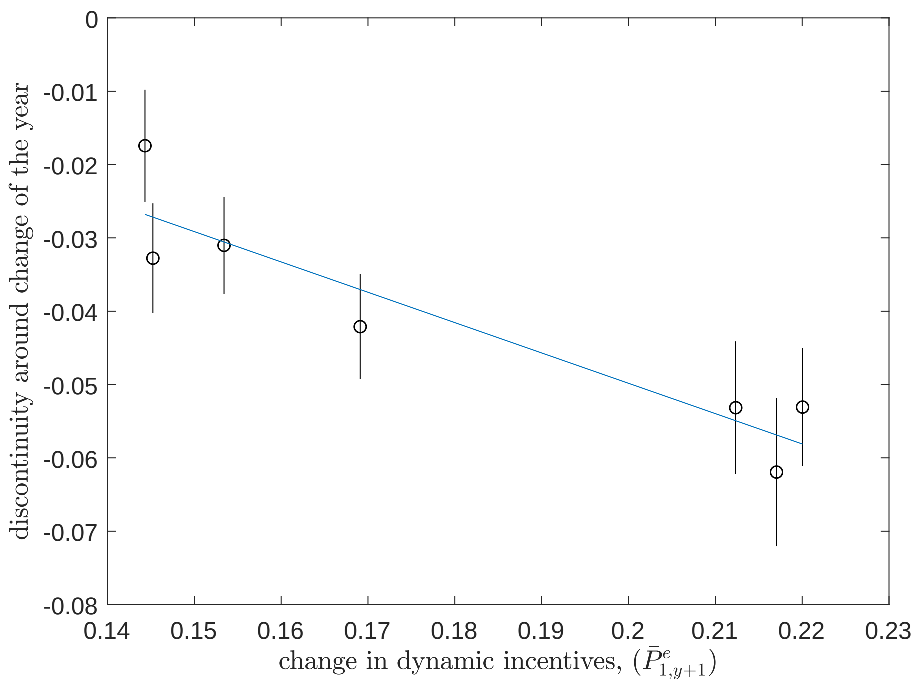
Notes: This figure replicates Figure 4 with the inclusion of care provided on weekends.

Figure A.25: Relationship between jump and $\bar{P}^e_{1,y+1}$ (PC at 5000)



Notes: This figure replicates Figure 4(a) when we pseudo-censor at €5000 instead of €500.

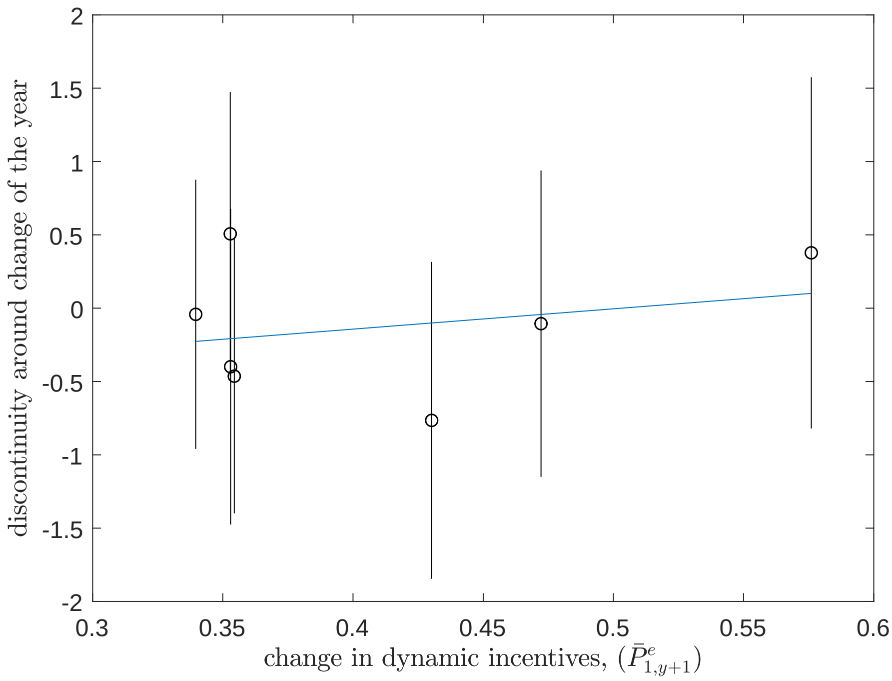
Figure A.26: Relationship between discontinuity size and $\bar{P}^e_{1,y+1}$ $(\ln(exp+1))$



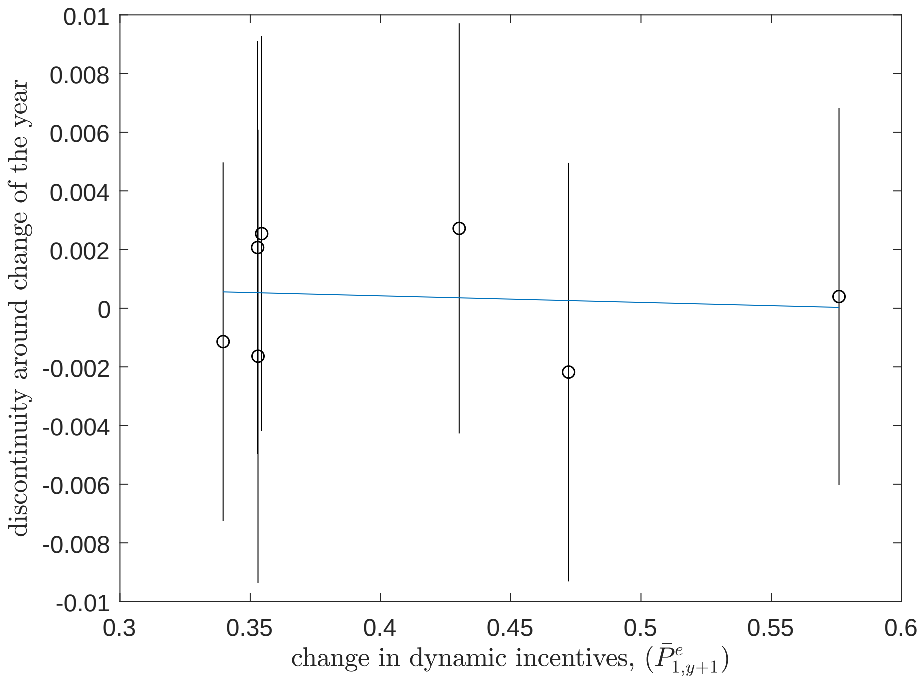
Notes: This figure replicates Figure 4(a) when we use the log of expenditure plus 1 as the dependent variable.

Figure A.27: Relationship between discontinuity sizes and dynamic incentives (placebo)

(a) Daily mean spending (pseudo-censored)

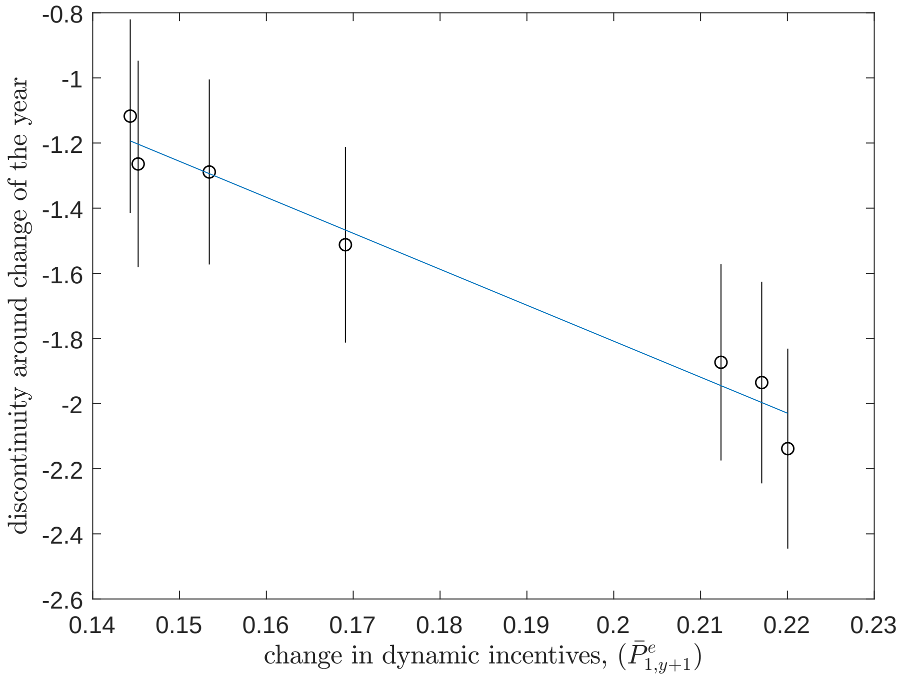


(b) Prob. any daily expenditure



Notes: This figure replicates Figure 4 for the placebo sample (individuals aged below 18).

Figure A.28: Removing last year-pair



Notes: This figure replicates Figure 4(a) when we remove the last year-pair from the analysis.