ORIGINAL PAPER

Ex-post moral hazard in the health insurance market: empirical evidence from German data

Stefanie Thönnes[1] (ID)

Abstract

In this paper, I analyze whether premium refunds can reduce ex-post moral hazard behavior in the health insurance market. I do so by estimating the effect of these refunds on different measures of medical demand. I use panel data from German sickness funds that cover the years 2006–2010 and I estimate effects for the year 2010. Applying regression adjusted matching, I find that choosing a tariff that contains a premium refund is associated with a significant reduction in the probability of visiting a general practitioner. Furthermore, the probability of visiting a doctor due to a trivial ailment such as a common cold is reduced. Effects are mainly driven by younger (and, therefore, healthier) individuals, and they are stronger for men than for women. Medical expenditures for doctor visits are also reduced. I conclude that there is evidence that premium refunds are associated with a reduction in ex-post moral hazard. Robustness checks support these findings. Yet, using observable characteristics for matching and regression, it is never possible to completely eliminate a potentially remaining selection bias and results may not be interpreted in a causal manner.

Keywords Health insurance · Ex-post moral hazard · Healthcare utilization · Premium refund

JEL Classification I11—Analysis of health care markets · I12—Health behavior · I13—Health insurance, public and private

Introduction

In many developed countries, health insurance systems suffer from increasing expenditures [1] which is a financing challenge. The increase in expenditures is mainly driven by technological progress, demographic change, and inefficiencies in the health care system, one of them being moral hazard [2].

In general, co-payment can be a device to reduce moral hazard. It may reduce demand for health care services by increasing the price paid by the consumer at the time of consumption [3, 4]. The magnitude of the effect depends on the price elasticity of demand. Imposed on price-elastic health care services, co-payment may be shown to reduce the demand.

Indeed, there is empirical evidence that co-payment—independent from the exact design—can reduce demand in the health insurance market. In the RAND Health Insurance Experiment, medical demand becomes smaller as the level of cost-sharing increases [5]. Similarly, in the Oregon Health Insurance Experiment, randomly extending insurance coverage increases the use of health care services [6].

Furthermore, there is non-experimental literature on compulsory co-payments. A rise in co-payments for doctor visits and drug prescriptions among retired public employees in the US reduces both kinds of medical utilization [7]. Likewise, for Switzerland, it has been found that in contrast to cost-sharing, full insurance coverage increases health care costs and decreases the probability of having zero health care expenditure [8]. In 2004, the German statutory health insurance (SHI) introduced a co-payment that had to be paid for every first doctor visit in each quarter. According to Farbmacher and Winter [9], it leads to a significant reduction in the probability of visiting a doctor by 4–8% points. Effects are higher for younger than for older adults. The authors find for young male adults that the co-payment reduces the number of doctor visits by 0.2–0.3 visits per quarter, while young women do not seem to be affected. In addition, in a subsequent study, Farbmacher et al. [10] estimate a significant

✉ Stefanie Thönnes
   thoennes@mail.uni-paderborn.de

1 Department of Economics, University of Paderborn,
   Warburger Straße 100, 33098 Paderborn, Germany

increase in the probability of not visiting a doctor, while the subgroup of elderly women with more severe diagnoses and a higher level of drug consumption is rarely affected. In contrast, Kunz and Winkelmann [11] use a different methodology and do not find any effect of this co-payment.

Finally, there is literature on settings in which individuals are free to choose more or less insurance coverage. Schmitz [2] concludes that with less insurance coverage, the probability of visiting a doctor is reduced for insured individuals with previously few doctor visits, while individuals with previously many doctor visits are not affected. At last, optional deductibles are found to reduce the number of doctor visits and the probability of visiting a specialist. Medical expenditures are decreased, also in the medium term. The effect is stronger for higher deductibles [12–14].

All in all, the literature suggests that deductibles—be they compulsory or optional—can reduce demand for health care services. Interestingly, the design of the cost-sharing scheme plays an important role. Hayen et al. [15] find that with a deductible scheme individuals react nearly twice as strongly compared to a premium refund scheme. However, the authors use data from The Netherlands, where insurance contracts are shaped differently from Germany. One crucial difference is that deductibles (premium refunds) increase (decrease) by one euro with every euro that is caused as costs to the sickness fund, until some threshold value is reached. Therefore, results cannot be generalized to the German system, where any health care resource use (apart from some exceptions) cuts the premium refund to zero (see below). Furthermore, the German system differs from the Dutch system with respect to the type of exceptions that do not cut the premium refund.

Premium refunds are rather new to the German SHI and will be analyzed in this paper. They work differently from deductibles but pursue the same goal. Besides being optional, they are known for not comprising any risk of loss for the insured. Broadly speaking, if individuals who chose a tariff including premium refunds do not make use of the health insurance for one year, they will be rewarded by cash. There is no risk of paying more than a person that stays in the default ("full insurance")[1] tariff, only the chance of missing the reward. Since individuals are often risk-averse, this may be a fitting solution when facing the incentives of deductibles without having the risk of paying more than in the default tariff. Insured persons are exempt from the premium refund as soon as they make use of their health insurance, so the reward scheme is highly nonlinear. Nonlinear

schemes are of high policy relevance and credible evidence on their impact was found earlier [10, 16, 17]. Yet, it could be a further improvement to look at specific diagnoses when analyzing moral hazard behavior.

My research question is whether premium refunds are a device to reduce ex-post moral hazard as well. Individuals that have opted for such a tariff will make use of the health insurance only if their utility from treatment is at least as high as their utility from forgoing treatment and receiving the premium. Due to information asymmetry, hospital treatment, drug prescriptions, and follow-up doctor visits are primarily decided on by the doctor. In contrast, the decision to visit a practitioner for the first time in the respective year often lies with the patient. This applies especially to general practitioners (GPs) and only to a lesser extent to specialists.

Moreover, one could imagine that participants of the premium refund tariff have an incentive to avoid visiting a doctor in the case of a trivial disease that can also be cured by means of self-medication, e.g., a common cold. Then, demand is supposed to be price-elastic. Assuming that nearly everybody gets a common cold once in a while, everybody is affected by this sickness in a similar way. If individuals visit a doctor due to a common cold, this does not mean that they are sicker than individuals with a common cold who do not visit a doctor. Instead, it reveals their behavior.

I use administrative data from two German sickness funds which both offer a premium refund tariff. I estimate the effect of choosing this tariff (in contrast to staying in the default tariff) in the year 2010 on the probability of visiting a GP or a specialist as well as on the probability of visiting a doctor due to a common cold in the same year. I am aware of the likely selection bias and use regression adjusted propensity score matching as well as a rich set of control variables. Selection is likely to result from the voluntary nature of the tariff. Younger and healthier individuals are especially more likely to opt for such an insurance scheme. Although multiple efforts are undertaken to reduce the selection bias, it is never possible to completely eliminate it using only matching on observables and OLS regression analysis. Therefore, results may not be interpreted in a causal manner. However, to support my findings, I carry out many robustness checks and I assess the level of remaining unobserved heterogeneity by applying a method proposed by Altonji et al. [18] and Oster [19]. Since this paper is motivated by increasing expenditures in the health care sector, I additionally estimate the effect on the sickness funds' medical expenditures for practitioners (GPs and specialists), although one has to keep in mind that due to information asymmetry in most cases, the patient cannot fully decide on the type and extent of treatment.

The literature suggests that the degree of moral hazard varies across individuals, e.g., by the extent of demand and the health status prior to the introduction (or removal) of any

---

[1] I call the default tariff in the German SHI "full insurance" although it exhibits some kinds of co-payment as well, e.g., for drug prescriptions or hospital stays. However, these are identical for individuals in both tariffs.

cost-sharing [2, 16], or by the type of disease the individual is suffering from [20]. Therefore, I repeat the analysis for subgroups to find out whether the effects are heterogeneous.

The contribution of this paper to the existing literature is first that by identifying relevant ICD-10-GM[2] diagnoses, it is possible to directly test whether a reduction in medical demand is due to reduced ex-post moral hazard. This is not always clear if medical expenditures or doctor visits are analyzed. Second, to the best of my knowledge, this is the first study that analyzes the effect of premium refunds among the German SHI on moral hazard behavior.

The paper is organized as follows. "Background" gives some background information on the legal regulations, while "Identification and estimation" discusses the identification and estimation strategy. In "Data", the data are described. Results are shown in "Results", and section "Discussion" discusses the results and concludes.

Background

The German health insurance system is characterized by the coexistence of a statutory and a private health insurance system. The vast majority of the population in Germany (85.34% in 2010) is covered by SHI [21, 22]. Everybody is obligated to insure themselves. However, not everyone can decide between the two systems. Only high earners, self-employed, and civil servants can (but do not have to) choose private health insurance. If they opt for the SHI, they are referred to as voluntarily insured, while the rest below pension age are mostly compulsorily insured. Within the system of the SHI, there were 169 sickness funds in 2010 [23]. Sickness funds have to accept all applicants, irrespective of their health status or income. The contribution to the SHI depends on the insurance member's gross wage only, not on the individual's health status. Dependent children and non-working spouses can be co-insured free of charge ("co-insured family members"). Insured persons are free to choose their favorite provider, regardless of the sickness fund that they are insured with. Most benefits[3] are identical for all sickness funds and are provided as benefits-in-kind. In addition, sickness funds can offer supplementary benefits which may vary between sickness funds but which represent only a small part of all medical services.

Premium refunds are rather new to the SHI. After some pilot projects [24], they were introduced to the whole system in 2007. In 2010, there were 152,571 individuals enrolled in the tariff [21] which corresponds to 0.2% of all statutorily insured persons. It can but does not have to be offered by the sickness fund. If the insurance company installs the tariff, it is offered to all members that have been insured with the sickness fund for at least three months. Enrollment in the tariff is voluntary and may start any time of the year. Insurance members decide whether they want to participate in the tariff, co-insured family members have to follow accordingly. If a person enrolls during the year, some sickness funds allow for retrospective enrollment (i.e., enrollment applies to the whole calendar year), while others restrict enrollment to the remainder of the year. Given that an insurance member and the co-insured family members do not cause expenditures within one calendar year, the insurance member receives a refund of 1/12 of his annual insurance contribution.[4] Otherwise, the refund is cut to zero. Thus, the plan is highly nonlinear. The only treatment allowed where the refund is not lost is for the under-18s, early diagnosis examinations, and prevention.[5]

If individuals choose the premium refund tariff, they are bound to the respective sickness fund for one year which is one reason why not everyone wants to enroll. Another reason for the relatively small share of enrollees is that sickness funds do not promote this tariff very strongly.[6] Furthermore, especially women in their childbearing years will not choose this tariff if they plan to collect a prescription for contraception every (second) quarter. Likewise, those under permanent medication do not have a reason to register for this tariff. Finally, for many individuals, it is more attractive to choose another tariff which is often offered by sickness funds besides the premium refund tariff, the so-called deductible tariff. Here, individuals pay a lower premium to their sickness fund, but in the event of a sickness, they bear the risk of paying more than the premium of the default tariff. The advantage of the deductible tariff is that even if individuals visit a doctor for a minor issue, they might only

---

[2] ICD-10-GM refers to "International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, German Modification".

[3] These benefits are recorded in the Code of Social Law Book V ("Sozialgesetzbuch V").

[4] Each sickness fund has its own contribution rate which applies to both individuals in the default tariff as well as those in the premium refund tariff, and which is identical for both. As a contribution, 7.3% of the employee's gross income is paid by the employer, the rest by the employee. For example, if the total contribution rate is 15.5% the insured pays 8.2%. If they earn 2000 euros gross per month, they pay a contribution of 164 euros each month. This is the amount a participant of the premium refund tariff may get back at the end of the year. For individuals in the default tariff, there is no premium refund at all.

[5] For individuals that have been less than 365 days insured, the refund tariff works analogously: These individuals paid less insurance contribution in this year (e.g., only for 11 months) but they can still get a premium refund of 1/12 of their annual insurance contribution, i.e., the absolute value of the potential premium refund is lower.

[6] Both sickness funds offer information on the premium refund tariff on their website. However, there was no additional information sent to the insured via newsletter.
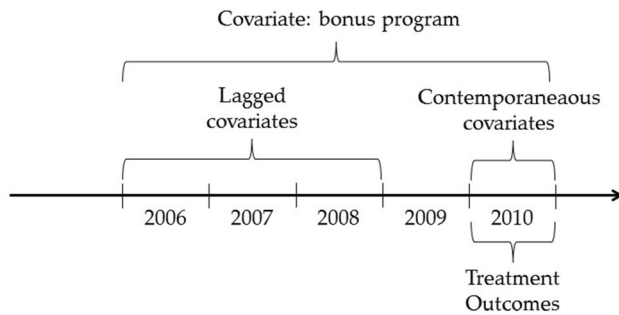
Fig. 1 Timeline. The figure shows in which years the treatment variable, outcomes, and the different covariates are measured

pay a small deductible, so that all in all, they are better off compared to the default tariff. Since insured individuals are not allowed to combine the premium refund tariff and the deductible tariff, many of those who are willing to choose any non-default tariff will select the deductible tariff.[7]

If sickness funds offer the premium refund tariff, they must prove to their respective supervision every three years that the tariff pays for itself, i.e., cross-subsidization is not allowed. This is supposed to prevent the rise of insurance contributions due to this tariff.

In addition, sickness funds are allowed to offer other tariffs. In the so-called bonus programs, the insured are rewarded if they can prove a healthy lifestyle and preventive actions. Again, participation is voluntary.

Identification and estimation

I observe two groups. The treatment group consists of individuals that participate in the premium refund tariff, whereas the control group does not. Treatment takes place during the whole year of 2010 (cf. Fig. 1). Outcomes, a group of various measures of medical demand, are quantified at the end of 2010 for the duration of the year, i.e., January 1, 2010, to December 31, 2010.

I am interested in the average treatment effect on the treated (ATT), which is the difference in demand for medical treatment between persons in the treatment group that have been treated on the one hand and persons in the treatment group had they not been treated on the other hand:

$$\tau_{\text{ATT}} = E[Y(1)|D = 1] - E[Y(0)|D = 1], \tag{1}$$

where $D = 1$ indicates that an individual belongs to the group that will choose the tariff. $Y(1)$ is the demand of

individuals that actually chose the tariff, whereas $Y(0)$ is the demand of these individuals had they not chosen it.

Naturally, the counterfactual $E[Y(0)|D = 1]$ is not known. Since participation is voluntary, individuals opting for the tariff differ from persons who do not, even if there was no treatment (selection bias). Therefore, I use the large group of non-participants to find individuals that are similar to the participants in all relevant (pre-treatment) characteristics.

The conditional independence assumption (CIA) will be violated if there are differences between participants and non-participants with respect to their risk, i.e., their health status, and with respect to their risk aversion [25]. Individuals maximize their expected utility. Therefore, if individuals expect high expenses in the future, they will prefer more coverage. The health status in the past is associated with the health status and the demand for health services, both in the future, and, therefore, also with the tariff choice. I use lagged values of health care claims and certain diagnoses as a proxy for the health status previous to the year 2010 (cf. Fig. 1, "Lagged covariates").[8]

Furthermore, risk-averse individuals might avoid co-payments of any kind and may, therefore, be less likely to choose the premium refund tariff. Simultaneously, they might show higher preventive effort [2]. As a proxy for the risk attitude toward health, I use the information of whether the individual participated in the sickness fund's bonus program in any of the years 2006–2010 (cf. Fig. 1, "Covariate: bonus program"). Individuals that participate in a bonus program are more concerned with their health status and are, therefore, suspected of being risk-averse. Also, Cutler et al. [25] assess the "receipt of preventive health care as a behavior that likely captures individual risk aversion". Since bonus programs primarily consist of preventive health activities, this is a useful proxy. I assume that risk attitude is stable over a period of five years, and that it is a valid signal if the individual participated in the bonus program in any of these years. The bonus program existed years before the premium refund tariff was implemented. Therefore, it is unlikely that choosing the premium refund tariff would affect participation in the bonus program. One limitation of this proxy is that the underlying risk aversion is presumably continuously distributed, while participation in the bonus program is a dichotomous measure. Therefore, it is useful to also match on pre-treatment outcomes (cf. Fig. 1, "Lagged covariates"). This approach accounts for historical factors that cause current differences in the dependent variable that are difficult

---

[7] In fact, in 2010 there were twice as many individuals enrolled in the deductible tariff compared to the premium refund tariff [21].

[8] To condition on lagged outcomes and on some lagged covariates, I use the average of the years 2006–2008. Observations of the year 2009 are not used as control variables because individuals could anticipate their participation in 2010 and choose to antedate medical demands. This behavior was shown by Chandra et al. [7].

to account for in other ways [26] and has frequently been used in applied research (see, e.g., [27–29]) to control for individual-specific unobserved heterogeneity. Being a rather unspecific approach, using lagged-dependent variables as proxy variables may also help to account for other sources of unobserved heterogeneity such as genetic factors or lifestyle factors.

I use socioeconomic information measured in 2010 to further adjust the two groups (cf. Fig. 1, "Contemporaneous covariates"). Besides age and gender, I control for insurance status and education as well as being insured with one of the two sickness funds.

While I chose some covariates due to economic theory, others were selected by an algorithm proposed by Imbens[9] [30]. In addition, the algorithm selected a large set of interaction terms which makes the model more flexible.

To take the CIA as given, it would be necessary to also consider the possibility of selection on moral hazard [31]. When opting for or against more insurance coverage,[10] individuals take into account how strongly they will react to an increase in insurance coverage. According to Finkelstein et al. [32], if deductibles are optional, those who are less responsive than average to consumer cost-sharing are more likely to choose deductibles. Individuals with higher price sensitivity would rather not choose deductibles. Although premium refunds and deductibles are not the same, they set similar incentives. Therefore, the estimated effects will resemble a lower bound and effects may even be up to two or three times higher [32] if participation was mandatory.

I use the Epanechnikov kernel estimator for the propensity score-based matching procedure.[11] Propensity score matching has the advantage of condensing the information of numerous matching variables into a one-dimensional measure. The Epanechnikov kernel estimator is appropriate for this application, because it takes many controls into consideration for every treated and gives more weight to rather similar than to rather different controls. Using a probit model, the propensity score is estimated as follows:

$$\text{participation} = \alpha_0 + X'\delta + u, \tag{2}$$

where X represents the vector of covariates (cf. Table 1, plus a long list of interaction terms), and u is the error term. As a result, treated individuals are matched to controls that have a similar but not identical propensity score. There may still be discrepancies between the covariates of the two groups, even though differences have already been reduced by the matching procedure. Hence, the estimator may still be biased. One can attempt to reduce this (residuary) bias using regression methods [33]. Therefore, I combine matching with regression adjustment.[12] Using the matched sample, I regress each outcome on participation in the premium refund tariff and on all control variables that have also been used for the matching procedure. The regression model is

$$Y = \beta_0 + \beta_1 \text{participation} + X'\gamma + \varepsilon, \tag{3}$$

where Y is one of the outcomes (cf. Table 1), X again represents the vector of covariates, and ε is the error term. In addition, the weights which result from the matching procedure are used in the regressions. In line with Schmitz and Westphal [34], in the OLS regressions, I employ robust standard errors, because they are easier to compute, even though they are slightly more conservative than bootstrapped standard errors [35].

The insured may be allowed to enroll retrospectively in the premium refund tariff, at the latest until the end of the calendar year. This leads to the problem that new participants are not necessarily affected by the tariff. Instead, one has to assume that a considerable share enrolls in the tariff by the end of the year if they discover they did not cause any insurance claims. I aim at removing this effect by eliminating all new participants of the year 2010 from the sample if they did not already participate in 2009.

The effect of more (or less) insurance coverage on medical demand consists of two parts [32]: The substitution effect is the moral hazard response, and therefore, the effect I am interested in. In addition, there may be an income effect, i.e., individuals with more insurance coverage can afford treatment which would be too expensive for them if they had less coverage. Here, the latter presumably does not exist. At the time of treatment, individuals paid the same insurance contribution as if they had stayed in the default tariff. In both tariffs, they have access to the same portfolio of benefits. The only difference is that at the end of the year, participants of the premium refund tariff lose a financial reward if they made demands for medical services. At the time of

---

[9] Imbens proposes pre-selecting variables that are assessed as being important according to economic theory. In addition, a set of variables is selected where it is not clear whether they should be included in the model. Each of these variables is tested by comparing a logistic regression of the treatment dummy on pre-selected variables with a logistic regression of the treatment dummy on pre-selected variables plus one of the variables that are to be tested. The variable where the likelihood ratio test statistic is the highest is included in the model. This procedure is repeated until the test statistic falls below some threshold. In line with Imbens [30], I use 1.00 as threshold value for linear terms and 2.71 for quadratic/interaction terms.

[10] With an increasing degree of cost-sharing, insurance coverage becomes smaller because the sickness fund only pays for a smaller part of the costs.

[11] For the estimation, I use the psmatch2 Stata command.

[12] Another advantage of combining propensity score matching and regression is that by comparing the distribution of propensity scores between the two groups participants with no or very few controls can be excluded from the analysis (i.e., the data are trimmed). Thereby, the two groups can be made more similar.

treatment, however, there should be no income effect. Therefore, what I will find is the substitution effect, i.e., the moral hazard response.

Whether the CIA is fulfilled cannot be directly tested. However, the assumption is supported if one does not find an effect of the treatment on a pseudo outcome, i.e., an outcome that is known to be unaffected by the treatment [33]. I repeat the analysis illustrated above, replacing outcomes with the pseudo outcome "probability of visiting a hospital in 2010". Treatment in hospital is mostly associated with severe illnesses. Therefore, the demand should be price-inelastic and the effect is expected to be zero.

Finally, to get a better idea of how strong the omitted variable bias may still be, I apply a method that was proposed by Altonji et al. [18] and further developed by Oster [19]. They had the idea that the degree of selection on observables is a guide to the degree of selection on unobservables. Using the Stata command psacalc, I estimate the treatment effect for the various outcomes under three different assumptions: selection on unobservables is half as big as/as big as/twice as big as selection on observables.

Data

The panel data cover the years 2006–2010 and result from the billing processes of two German sickness funds. They cover the annual costs per insurance member, including co-insured family members but excluding under-18s. Costs contain expenditures for hospitalization, doctor visits, drugs, sickness payments, as well as so-called other costs.[13] Thereby, all relevant fields that are covered by the SHI, except for information on visits to the dentist, are included. Annual costs (and count variables, e.g., the number of doctor visits) are standardized (averaged) according to the number of members of the specific family[14] as well as the number of days the family was insured with this fund in the respective year. The sample is limited to individuals who were insured for at least 150 days in the year 2010 as well as 150 days in sum of the years 2006–2008. This was done, because observing individuals for a few days only may lead

to biased results. Furthermore, participants of the year 2010 that had not participated in the 2009 tariff were excluded from the sample.[15]

Beyond costs, information on the date and the ICD-10-GM diagnosis for any contact with the health care system is available. To identify doctor visits due to a common cold, I use two different measures—the ICD-10-GM codes J00 (acute rhinopharyngitis) and J00–J06 (acute infection of the upper respiratory system). I identify treatment of the common cold in the data on practitioners and at hospitals' outpatient departments. The data on practitioners differentiate between GPs and specialists.[16] For indicator variables (e.g., on diagnoses), the maximum per family is considered. Moreover, some socioeconomic information on the insured person is available. Finally, it is known whether the person participated in the bonus program. Table 1 provides an overview of all variables used in this paper and explains what they measure.

All in all, the insurance members' structure in the sample is similar to that in the SHI with respect to gender and age (cf. Table 9 in the "Appendix"). For 2010, the raw sample contains 751,687 insurance members. After applying the above-mentioned inclusion criteria, the sample contains 439,143 insurance members, whereof 13,187 participated in the premium refund tariff. Thereof, 1492 members received a premium refund.[17] Once individuals chose the tariff, they often stayed with it for many years. Of the 13,187 participants in 2010, 12,120 and 10,072 individuals had already participated in 2008 and 2007, respectively.

This study analyzes the effect of premium refunds on a variety of outcomes. Table 2 shows mean values and standard deviations and reveals how these outcomes are influenced by the data processing. It is noticeable that individuals that participate in the premium refund tariff have lower medical demand compared to non-participants with respect to nearly all measures. Furthermore, it can be seen that trimming the data (column 2 vs. column 1) primarily affects the treatment group, while matching (column 3 vs. column 2) mainly has an influence on the control group.

---

[13] "Other costs" comprise all remaining types of costs and are therefore very heterogeneous. They contain payments for rehabilitation, prevention courses, and home healthcare products, amongst others. When constructing the sum of "other costs" that occur in the sickness fund, I exclude premiums that were paid out to the insured for participating in the premium refund tariff or in a bonus program.

[14] Here, a family consists of the insurance member and his or her co-insured family members. If there are no co-insured family members, the data remain at the individual level. Otherwise, they are condensed to the family level. If both spouses are insurance members (instead of one being the insurance member and one being co-insured), they are recorded separately in the data and not as one family.

[15] See section "Identification and estimation" for the motivation of this approach.

[16] This information has only been available since July 2008 and can therefore not be used for the matching process.

[17] Among participants who received a premium refund, the average value of the premium refund is 294.44 euros (minimum 22.38 euros, at most 744.60 euros).

Table 1 Variable description

Variable	Definition/further description	Measured in year(s)
Dependent variables (outcomes)		
Doctor visit	= 1 if at least one doctor visit	2010
Doctor visit (GP)	= 1 if at least one visit to the GP	2010
Doctor visit (specialist)	= 1 if at least one visit to a specialist	2010
No. of doctor visits	Number of doctor visits	2010
No. of doctor visits (GP)	Number of visits to the GP	2010
No. of doctor visits (specialist)	Number of visits to a specialist	2010
Common cold (J00)	= 1 if at least one diagnosis J00	2010
Common cold (J00–J06)	= 1 if at least one diagnosis of J00 to J06	2010
Expenditures on doctor visits	Expenditures on doctor visits, measured in euros	2010
Expenditures on doctor visits (GP)	Expenditures on visits to the GP, measured in euros	2010
Expenditures on doctor visits (specialist)	Expenditures on visits to a specialist, measured in euros	2010
Explanatory variables		
Age	Measured in years	2010
Male	= 1 if male	2010
Bonus program in 2006–2010	= 1 if participated in bonus program in any of the years 2006–2010	2006–2010
Sickness fund 2	= 1 if individual is a member of sickness fund 2	2010
Compulsorily insured	= 1 if individual is compulsorily insured	2010
Voluntarily insured	= 1 if individual is voluntarily insured	2010
Unemployed	= 1 if individual is unemployed	2010
Pensioner	= 1 if individual is a pensioner	2010
Educational category 1	= 1 if no apprenticeship and no university-entrance diploma	2010
Educational category 2	= 1 if apprenticeship but no university-entrance diploma	2010
Educational category 3	= 1 if no apprenticeship but university-entrance diploma	2010
Educational category 4	= 1 if apprenticeship and university-entrance diploma	2010
Educational category 5	= 1 if degree of university of applied science	2010
Educational category 6	= 1 if university degree	2010
Positive expenditures total, before	= 1 if total expenditures > 0 euro in any of the years 2006–2008	2006–2008
Expenditures total, before	Measured in euros; average per year	2006–2008
Doctor visit, before	= 1 if at least one doctor visit in any of the years 2006–2008	2006–2008
Expenditures on doctor visits, before	Measured in euros; average per year	2006–2008
No. of doctor visits, before	Average per year	2006–2008
Hospital visit, before	= 1 if at least one hospital visit in any of the years 2006–2008	2006–2008
Expenditures on hospital visits, before	Measured in euros; average per year	2006–2008
Drug prescription, before	= 1 if at least one drug prescription in any of the years 2006–2008	2006–2008
No. of drug prescriptions, before	Average per year	2006–2008
Sickness absence, before	= 1 if at least one sickness absence in any of the years 2006–2008	2006–2008
Expenditures on sickness benefit, before	Measured in euros; average per year	2006–2008
No. of times on sickness absence, before	Average per year	2006–2008
No. of days on sickness absence, before	Average per year	2006–2008
Other costs, before	Measured in euros; average per year	2006–2008
Common cold (J00), before	= 1 if at least one diagnosis J00 in any of the years 2006–2008	2006–2008
Common cold (J00–J06), before	= 1 if at least one diagnosis of J00 to J06 in any of the years 2006–2008	2006–2008
Pseudo outcome		
Hospital visit (2010)	= 1 if at least one hospital visit during 2010	2010

GP general practitioner, *No.* number

Table 2 Descriptive statistics: mean values of outcomes in 2010

Outcome	(1)		(2)		(3)	
	Untrimmed sample		Trimmed sample		Trimmed and matched sample	
	Treatment group	Control group	Treatment group	Control group	Treatment group	Control group
Doctor visit	0.85 (0.35)	0.89 (0.31)	0.87 (0.34)	0.89 (0.31)	0.87 (0.34)	0.88 (0.32)
Doctor visit (GP)	0.76 (0.43)	0.82 (0.39)	0.78 (0.41)	0.82 (0.39)	0.78 (0.41)	0.81 (0.39)
Doctor visit (specialist)	0.72 (0.45)	0.76 (0.43)	0.73 (0.44)	0.76 (0.43)	0.73 (0.44)	0.74 (0.44)
No. of doctor visits	8.23 (9.07)	12.24 (14.29)	9.39 (10.15)	11.91 (13.38)	9.39 (10.15)	9.88 (10.98)
No. of doctor visits (GP)	3.40 (3.85)	5.16 (6.57)	3.86 (4.28)	4.99 (6.01)	3.86 (4.28)	4.16 (4.72)
No. of doctor visits (specialist)	4.45 (6.42)	6.61 (9.80)	5.15 (7.23)	6.50 (9.31)	5.14 (7.23)	5.32 (7.73)
Common cold (J00)	0.01 (0.12)	0.02 (0.15)	0.02 (0.13)	0.02 (0.16)	0.02 (0.13)	0.02 (0.15)
Common cold (J00–J06)	0.22 (0.42)	0.28 (0.45)	0.26 (0.44)	0.28 (0.45)	0.26 (0.44)	0.28 (0.45)
Expenditures on doctor visits	241.87 (314.96)	352.41 (633.88)	275.02 (352.67)	341.29 (512.27)	275.02 (352.67)	288.16 (418.64)
Expenditures on doctor visits (GP)	90.49 (96.16)	128.49 (159.66)	100.07 (104.74)	124.94 (155.97)	100.07 (104.74)	107.23 (118.83)
Expenditures on doctor visits (specialist)	147.13 (266.00)	214.91 (522.78)	170.00 (299.62)	208.98 (430.65)	170.00 (299.62)	174.70 (338.67)
N	13,091	380,996	5090	360,251	5090	5069

Standard deviation in parentheses. The table shows values before and after trimming the data as well as after matching the two groups. Due to missings in the data, number of doctor visits of GP and specialist do not add to the overall number. The same applies to expenditures on doctor visits and (in a modified manner) to the share of individuals with at least one doctor visit ("Doctor visit", first panel)

GP general practitioner, *No.* number

Results

Matching quality

After trimming the data, participants of the tariff still differ in some dimensions from unmatched non-participants (cf. Table 3). This becomes obvious through the standardized bias which lies far above 5% for most of the variables. Both groups are nearly of the same age and have a similar probability of participating in the bonus program. The distribution of the insurance status and education is similar between the two groups. This also holds for all probabilities of medical utilization (e.g., the probability of visiting a doctor due to a common cold). However, on average, the share of men is higher in the treatment group. Non-participants, on average, cause higher costs. This holds for all kinds of costs. Moreover, the number of times that they make use of the health care system (e.g., the number of drug prescriptions) is higher than for participants.

After the matching procedure, the average value of all covariates has converged between the treatment and the matched control group (cf. Table 3). The standardized bias is less than 5% for all variables that were used for matching. Thus, the matching procedure is successful [36]. Common support exists after having carried out trimming procedures.[18]

Estimation results

Estimation results are presented in Table 4.[19] The probability of visiting a GP is significantly reduced by 2.6% points. In contrast, the effect on the probability of visiting a specialist is smaller and only marginally significant. These findings are in line with theory. Likewise, the number of visits to the GP is significantly reduced by 0.3 visits (−7.4%), while there is only a smaller reduction of visits to a specialist (−0.2 visits or −3.5%, respectively). Moreover, I find that participants have a 0.7 or 2.1% point lower probability of visiting a doctor due to a common cold (depending on the definition of the ailment). This is a further indication that ex-post moral hazard behavior has been reduced. As expected, individuals avoid visiting a doctor due to trivial ailment such as a

Footnote 18 (continued)

group. Therefore, the data had to be trimmed relatively strongly. Figure 2 suggests trimming the data at propensity score = 0.1 because there are few observations in the control group with a propensity score higher than 0.1. Thereby, 8001 participants and 20,745 non-participants were excluded from the analysis. Since other authors may have decided differently on the trimming threshold, I will vary this threshold in the robustness checks.

[19] Table 4 only shows the coefficients of interest, i.e., the treatment coefficient of each of the regressions run. Additional results are shown in the "Appendix" in Table 10 (linear coefficients of the propensity score estimation) and Table 11 (linear coefficients of the various regressions run in the main specification).

---

[18] Treatment and control group were rather different with respect to the propensity score. Figure 2 in the "Appendix" shows the distribution of the untrimmed propensity score for the treatment and control

Table 3 Descriptive statistics: mean values of covariates

Covariate	Treatment group	Unmatched control group		Matched control group	
	Mean	Mean	S.B.	Mean	S.B.
Age	41.86 (11.40)	41.40 (12.32)	2.4	41.57 (11.36)	0.9
Male	0.71 (0.46)	0.59 (0.49)	25.1	0.70 (0.46)	0.3
Bonus program in 2006–2010	0.07 (0.26)	0.06 (0.24)	4.4	0.06 (0.25)	3.2
Sickness fund 2	0.12 (0.35)	0.72 (0.45)	− 153.4	0.13 (0.34)	− 2.1
Compulsorily insured	0.89 (0.31)	0.86 (0.35)	8.6	0.87 (0.33)	4.6
Voluntarily insured	0.08 (0.27)	0.08 (0.27)	− 1.0	0.09 (0.28)	− 4.0
Unemployed	0.00 (0.04)	0.02 (0.14)	− 17.3	0.00 (0.06)	− 1.1
Pensioner	0.02 (0.14)	0.03 (0.17)	− 6.4	0.02 (0.14)	− 0.0
Educational category 1	0.06 (0.24)	0.11 (0.32)	− 17.6	0.07 (0.25)	− 2.4
Educational category 2	0.71 (0.46)	0.63 (0.48)	16.6	0.71 (0.46)	− 0.2
Educational category 3	0.05 (0.21)	0.07 (0.25)	− 9.3	0.04 (0.20)	2.1
Educational category 4	0.08 (0.26)	0.08 (0.26)	− 0.1	0.07 (0.26)	0.7
Educational category 5	0.04 (0.21)	0.04 (0.19)	2.7	0.04 (0.21)	− 0.1
Educational category 6	0.06 (0.24)	0.08 (0.26)	− 4.6	0.06 (0.24)	0.4
Positive expenditures total, before	0.94 (0.24)	0.93 (0.25)	3.8	0.95 (0.22)	− 2.3
Expenditures total, before	492.79 (1103.19)	694.63 (1536.11)	− 15.1	494.41 (1149.67)	− 0.1
Doctor visit, before	0.94 (0.24)	0.93 (0.26)	5.0	0.95 (0.23)	− 2.9
Expenditures on doctor visits, before	194.63 (212.99)	246.00 (281.47)	− 20.6	194.33 (211.76)	0.1
No. of doctor visits, before	7.50 (7.01)	9.53 (8.94)	− 25.3	7.55 (6.92)	− 0.7
Hospital visit, before	0.20 (0.40)	0.23 (0.42)	− 8.3	0.20 (0.40)	0.5
Expenditures on hospital visits, before	170.79 (606.26)	246.62 (828.58)	− 10.4	173.52 (614.02)	− 0.4
Drug prescription, before	0.85 (0.35)	0.85 (0.35)	− 0.2	0.85 (0.35)	− 0.2
No. of drug prescriptions, before	2.56 (3.44)	3.66 (4.65)	− 27.0	2.58 (3.27)	− 0.6
Sickness absence, before	0.72 (0.45)	0.69 (0.46)	6.9	0.72 (0.45)	0.3
Expenditures on sickness benefit, before	59.03 (426.77)	81.72 (488.07)	− 4.7	57.06 (402.95)	0.4
No. of times on sickness absence, before	0.77 (0.90)	0.91 (1.16)	− 13.2	0.78 (0.90)	− 0.9
No. of days on sickness absence, before	7.79 (15.67)	10.23 (20.47)	− 13.4	7.87 (14.93)	− 0.4
Other costs, before	113.82 (260.01)	95.53 (322.54)	6.2	107.19 (298.64)	2.3
Common cold (J00), before	0.05 (0.21)	0.05 (0.22)	− 2.0	0.04 (0.21)	1.0
Common cold (J00–J06), before	0.48 (0.50)	0.50 (0.50)	− 3.3	0.48 (0.50)	− 0.3
N	5090	360,251		5069	

Standard deviation in parentheses. Educational category 1: No apprenticeship, no university-entrance diploma. Educational category 2: Apprenticeship, no university-entrance diploma. Educational category 3: No apprenticeship, university-entrance diploma. Educational category 4: Apprenticeship, university-entrance diploma. Educational category 5: Degree of university of applied science. Educational category 6: University degree. The data presented have already been trimmed. In addition to the variables presented here, numerous quadratic and interaction terms are used as covariates. Covariates concerning diagnoses, costs, or other measures of medical utilization are measured during the years 2006–2008. Socioeconomic covariates are measured in 2010

S.B. standardized bias, *No.* number

common cold. With a magnitude of 8 or 35%, respectively, this reduction is substantial.

Furthermore, I find a significant reduction in the medical expenditures for visits to the GP of 7 euros, while there is no significant reduction in expenditures for specialists. Although 7 euros does not seem to be much, it corresponds to a decrease of 7%, which is substantial.

Sensitivity analysis

I test whether results are stable and I carry out numerous robustness checks. First of all, to find support so that the CIA is fulfilled, I run regressions for the pseudo outcome (cf. Table 5). As expected, there is no significant effect of participating in the premium refund tariff in 2010 on the probability of visiting a hospital in the same year, and the coefficient is close to zero. Since the CIA cannot be directly

tested, this is not a proof, but it supports the assumption. It implies that the treated observations are not distinct from the controls in that the distribution of Y(0) for the treated units is comparable to the distribution of Y(0) for the controls.

Next, I vary the trimming procedure, since there is some area of discretion. For most outcomes, this does not lead to considerable differences (cf. Table 6, columns 1 and 2) and results are qualitatively robust to the exact cutoff for the trimming procedure, although they tend to become slightly smaller. It is noticeable that the probability of visiting a specialist becomes insignificant, and an effect should not be assumed. Furthermore, I vary the bandwidth from kernel matching. Exemplarily, results are shown for a bandwidth of 0.01 (cf. column 3). They are essentially the same as those in the main specification.

I also try other matching estimators that rely on the propensity score. For nearest neighbor matching (1:30, cf.

Table 4 Estimation results (main specification)

Outcome	Kernel matching plus regression
Doctor visit	− 0.018*** (0.004)
Doctor visit (GP)	− 0.026*** (0.005)
Doctor visit (specialist)	− 0.011* (0.006)
No. of doctor visits	− 0.479*** (0.121)
No. of doctor visits (GP)	− 0.284*** (0.051)
No. of doctor visits (specialist)	− 0.181** (0.091)
Common cold (J00)	− 0.007*** (0.002)
Common cold (J00–J06)	− 0.021*** (0.006)
Expenditures on doctor visits	− 13.227*** (4.443)
Expenditures on doctor visits (GP)	− 6.883*** (1.236)
Expenditures on doctor visits (specialist)	− 5.060 (3.920)
N treated	5090
N controls	360,251
N controls (weighted)	5069

Robust standard errors in parentheses. ***$p<0.01$, **$p<0.05$, *$p<0.10$. Each line contains the coefficient of treatment in separate regressions, i.e., for different outcomes. Bandwidth = 0.005, trimming at 0.1

GP general practitioner, *No.* number

Table 5 Robustness checks I

Outcome	Kernel matching plus regression
Hospital visit (2010)	− 0.006 (0.004)
N treated	5090
N controls	360,251
N controls (weighted)	5069

Robust standard errors in parentheses. ***$p<0.01$, **$p<0.05$, *$p<0.10$. The table contains the coefficient of treatment for the pseudo-outcome. Bandwidth = 0.005, trimming at 0.1. Covariates are measured in the same years as in the main specification

column 4), results are qualitatively similar to those in the main specification, only slightly smaller. For radius matching combined with regression, results are virtually the same as those in the main specification (cf. column 5). Moreover, I extend the minimum days an individual can be observed in the data from 150 to 365 (cf. column 6). This does not affect the results. Subsequently, instead of pooling the years 2006–2008 to create the lagged covariates and instead of leaving out 2009, I match treated and controls in the years 2006–2009 separately (cf. column 7). It is noticeable that the results are qualitatively the same as in the main specification, even if slightly smaller. Finally, I refrain from matching and trimming the data. Instead, I use OLS. The advantage of matching the two groups and trimming the data is that the common support can be ensured and groups can be made more similar. However, I want to assess its effect on the estimation. OLS results (cf. column 8) are weaker in magnitude but qualitatively similar to the main specification. All in all, results are stable over this variety of robustness checks.

Finally, to get a better idea of how strong the omitted variable bias still is, I apply Oster's method [19] as already described above. Columns 1 and 2 in Table 7 show that, for the trimmed and matched data, the use of control variables in the regression is not important. Columns 3–5 make different assumptions concerning the degree of selection on unobservables relative to selection on observables. It becomes obvious that no matter whether selection on unobservables is smaller than (column 3), equal to (column 4), or bigger (column 5) than selection on observables, results are very stable, which is another reassuring result indicating that selection on unobservables is not strong in this application.

Effect heterogeneity

Furthermore, I analyze how the effects are composed, i.e., whether subgroups are affected differently. I differentiate individuals by gender and age group. According to Table 8 (column 1), the subgroup of men reacts more strongly to the tariff's incentives than the whole sample. For women (column 2), it is noticeable that I do not find a significant negative effect on the probability of visiting a GP. Some effects found in the overall sample become insignificant for women. Obviously, men react stronger to the premium refund tariff's incentives than women.

Although there are three age groups that were analyzed, results show that they could be condensed into two groups. Individuals aged 34 and younger[20] (column 3) and those aged 35–49 (column 4) react very similarly. Results are qualitatively the same as in the whole sample, but effects are slightly stronger. In contrast, individuals aged 50 and

---

[20] The youngest participant is 20 years old.

Table 6 Robustness checks II

Outcome	(1) Kernel matching plus regression (modified trimming 1)	(2) Kernel matching plus regression (modified trimming 2)	(3) Kernel matching plus regression (modified bandwidth)	(4) Nearest neighbor matching (1:30)
Doctor visit	−0.017*** (0.004)	−0.013*** (0.004)	−0.018*** (0.004)	−0.011** (0.005)
Doctor visit (GP)	−0.024*** (0.005)	−0.021*** (0.005)	−0.027*** (0.005)	−0.019*** (0.006)
Doctor visit (specialist)	−0.007 (0.006)	−0.006 (0.005)	−0.011* (0.006)	−0.001 (0.007)
No. of doctor visits	−0.450*** (0.112)	−0.351*** (0.110)	−0.487*** (0.120)	−0.327** (0.150)
No. of doctor visits (GP)	−0.238*** (0.049)	−0.222*** (0.047)	−0.288*** (0.051)	−0.242*** (0.063)
No. of doctor visits (specialist)	−0.202** (0.083)	−0.123 (0.083)	−0.186** (0.091)	−0.079 (0.107)
Common cold (J00)	−0.007*** (0.002)	−0.007*** (0.002)	−0.007*** (0.002)	−0.007*** (0.002)
Common cold (J00–J06)	−0.020*** (0.006)	−0.016*** (0.005)	−0.021*** (0.006)	−0.017*** (0.006)
Expenditures on doctor visits	−12.587*** (4.153)	−9.327** (4.083)	−13.404*** (4.437)	−7.899 (5.292)
Expenditures on doctor visits (GP)	−5.694*** (1.180)	−5.199*** (1.139)	−6.935*** (1.238)	−5.728*** (1.562)
Expenditures on doctor visits (specialist)	−5.343 (3.604)	−2.327 (3.553)	−5.147 (3.915)	−1.369 (4.482)
N treated	5787	6205	5090	5090
N controls	366,246	369,199	360,251	360,251
N controls (weighted)	5768	6197	5069	–

Outcome	(5) Radius matching plus regression	(6) Kernel matching plus regression (at least 365 days insured)	(7) Kernel matching plus regression (lagged covariates not pooled)	(8) OLS
Doctor visit	−0.018*** (0.004)	−0.018*** (0.004)	−0.015*** (0.004)	−0.003 (0.003)
Doctor visit (GP)	−0.027*** (0.005)	−0.026*** (0.005)	−0.025*** (0.005)	−0.015*** (0.004)
Doctor visit (specialist)	−0.011* (0.006)	−0.010* (0.006)	−0.006 (0.006)	−0.005 (0.004)
No. of doctor visits	−0.484*** (0.121)	−0.436*** (0.122)	−0.342*** (0.113)	−0.358*** (0.078)
No. of doctor visits (GP)	−0.287*** (0.051)	−0.275*** (0.052)	−0.198*** (0.050)	−0.212*** (0.036)
No. of doctor visits (specialist)	−0.184** (0.091)	−0.149 (0.092)	−0.141 (0.087)	−0.163*** (0.059)
Common cold (J00)	−0.007*** (0.002)	−0.007*** (0.002)	−0.006*** (0.002)	−0.005*** (0.001)
Common cold (J00–J06)	−0.021*** (0.006)	−0.021*** (0.006)	−0.014** (0.006)	−0.017*** (0.004)
Expenditures on doctor visits	−13.322*** (4.446)	−11.641** (4.496)	−10.751** (4.375)	−3.776 (3.408)
Expenditures on doctor visits (GP)	−9.911*** (1.237)	−6.841*** (1.233)	−4.754*** (1.229)	−8.087*** (0.990)
Expenditures on doctor visits (specialist)	−5.121 (3.923)	−3.528 (3.980)	−4.297 (3.915)	1.365 (3.078)
N treated	5090	5019	4585	13,091
N controls	360,251	332,481	317,522	380,996
N controls (weighted)	5090	5000	4563	–

Robust standard errors in parentheses. ***$p<0.01$, **$p<0.05$, *$p<0.10$. Each line contains the coefficient of treatment in separate regressions, i.e., for different outcomes. Ad (1): similar to the main specification but modified trimming; bandwidth = 0.005, trimming at 0.12. Ad (2): similar to the main specification but modified trimming; bandwidth = 0.005, trimming at 0.14. Ad (3): similar to the main specification but modified bandwidth; bandwidth = 0.01, trimming at 0.1. Ad (4): nearest neighbor matching based on the propensity score. Ad (5): radius = 0.005, trimming at 0.1. Ad (6): bandwidth = 0.005, trimming at 0.1. Instead of limiting the sample to individuals that are at least 150 days insured in 2010 and in 2006–2008, here, this threshold is extended to 365 days. Ad (7): bandwidth = 0.01, trimming at 0.1. Instead of using the average of 2006–2008 for lagged covariates, here, lagged covariates are measured in 2006, 2007, 2008, and 2009 separately. Ad (8): data are untrimmed. Regressions are unweighted

GP general practitioner, *No.* number

Table 7 Robustness checks III

Outcome	(1) Uncontrolled	(2) Controlled	(3) Tests according to Oster [19]	(4)	(5)
			$\text{Selection}_{\text{Unobs}} < \text{Selection}_{\text{Obs}}$	$\text{Selection}_{\text{Unobs}} = \text{Selection}_{\text{Obs}}$	$\text{Selection}_{\text{Unobs}} > \text{Selection}_{\text{Obs}}$
Doctor visit	− 0.018	− 0.018	− 0.017	− 0.017	− 0.017
Doctor visit (GP)	− 0.027	− 0.026	− 0.026	− 0.026	− 0.026
Doctor visit (specialist)	− 0.009	− 0.011	− 0.013	− 0.014	− 0.016
No. of doctor visits	− 0.488	− 0.479	− 0.470	− 0.463	− 0.453
No. of doctor visits (GP)	− 0.301	− 0.284	− 0.267	− 0.254	− 0.236
No. of doctor visits (specialist)	− 0.177	− 0.181	− 0.188	− 0.193	− 0.199
Common cold (J00)	− 0.007	− 0.007	− 0.008	− 0.008	− 0.008
Common cold (J00–J06)	− 0.021	− 0.021	− 0.020	− 0.020	− 0.020
Expenditures on doctor visits	− 13.141	− 13.227	− 13.353	− 13.439	− 13.548
Expenditures on doctor visits (GP)	− 7.162	− 6.883	− 6.594	− 6.374	− 6.061
Expenditures on doctor visits (specialist)	− 4.706	− 5.060	− 5.756	− 6.167	− 6.632
N treated	5090				
N controls	360,251				
N controls (weighted)	5069				

Robust standard errors in parentheses. Each line contains the coefficient of treatment in separate regressions, i.e., for different outcomes. Significance is not shown in this table. Ad (1): main specification without control variables. Ad (2): main specification. Ad (3) to (5): tests according to Oster [19]. The coefficients shown are estimated using different assumptions. Ad (3): assumption that selection on unobservables is half as big as selection on observables. Ad (4): assumption that selection on unobservables is as big as selection on observables. Ad (5): assumption that selection on unobservables is twice as big as selection on observables

GP general practitioner, *No.* number

Table 8 Estimation results by subgroups

Outcome	(1) Men	(2) Women	(3) Aged 34 and younger	(4) Aged 35–49	(5) Aged 50 and older
Doctor visit	− 0.032*** (0.006)	0.015*** (0.006)	− 0.022** (0.009)	− 0.018*** (0.007)	− 0.005 (0.007)
Doctor visit (GP)	− 0.032*** (0.007)	− 0.011 (0.009)	− 0.038*** (0.011)	− 0.025*** (0.008)	− 0.004 (0.008)
Doctor visit (specialist)	− 0.027*** (0.008)	0.024*** (0.008)	− 0.024** (0.011)	− 0.010 (0.009)	0.005 (0.010)
No. of doctor visits	− 0.528*** (0.133)	− 0.299 (0.257)	− 0.608*** (0.176)	− 0.589*** (0.183)	− 0.012 (0.275)
No. of doctor visits (GP)	− 0.244*** (0.061)	− 0.285*** (0.096)	− 0.362*** (0.066)	− 0.283*** (0.078)	− 0.004 (0.127)
No. of doctor visits (specialist)	− 0.271*** (0.094)	− 0.025 (0.207)	− 0.240* (0.145)	− 0.266* (0.137)	− 0.037 (0.189)
Common cold (J00)	− 0.008*** (0.002)	− 0.005 (0.004)	− 0.006* (0.004)	− 0.009*** (0.003)	− 0.004 (0.003)
Common cold (J00–J06)	− 0.019*** (0.007)	− 0.026** (0.011)	− 0.027** (0.011)	− 0.020** (0.009)	− 0.015 (0.011)
Expenditures on doctor visits	− 15.640*** (4.793)	− 9.219 (9.622)	− 19.567*** (6.068)	− 13.786** (6.932)	1.457 (10.213)
Expenditures on doctor visits (GP)	− 6.697*** (1.463)	− 6.745*** (2.336)	− 8.485*** (1.575)	− 7.464*** (1.879)	− 0.386 (2.980)
Expenditures on doctor visits (specialist)	− 7.587* (4.085)	− 1.947 (8.817)	− 9.792* (5.478)	− 4.857 (6.129)	1.909 (8.875)
N treated	3551	1479	1550	2137	1323
N controls	210,197	144,134	118,007	129,920	86,176
N controls (weighted)	3540	1476	1545	2132	1306

Robust standard errors in parentheses. ***$p<0.01$, **$p<0.05$, *$p<0.10$. Each line contains the coefficient of treatment in separate regressions, i.e., for different outcomes. Ad (1): bandwidth = 0.005, trimming at 0.1. Ad (2): bandwidth = 0.001, trimming at 0.1, standardized bias slightly too high (7.9 at most). Ad (3): bandwidth = 0.005, trimming at 0.1. Ad (4): bandwidth = 0.01, trimming at 0.1. Ad (5): bandwidth = 0.005, trimming at 0.1

GP general practitioner, *No.* number

older[21] (column 5) do not react to participation in the tariff at all. All coefficients are insignificant and they are mostly close to zero.

Discussion

This paper examines whether premium refunds are a suitable instrument to reduce ex-post moral hazard in the health insurance market. I use panel data covering the years 2006–2010 which result from the billing processes of two German sickness funds. I analyze the effect of participating in the premium refund tariff in 2010 on several health measures in the same year by combining propensity score matching and regression.

I find that participating in the premium refund tariff is associated with a significant reduction in the probability of visiting a GP (-2.6% points). This is in contrast to Felder and Werblow [12], but in line with Farbmacher and Winter [9] and Health Policy Brief [6], although they report a higher reduction. However, this is not unexpected: since potential selection on moral hazard [31] was not accounted for in the present study, the estimated effects in this paper resemble a lower bound [32] and true effects may be higher. Like Farbmacher and Winter [9], I also find that effects are higher for younger than for older individuals and that men are more strongly affected than women. The number by which doctor visits are reduced in this study is of a similar magnitude as in Farbmacher and Winter [9] and the effect goes in the same direction as in Chandra et al. [7]. In addition, I find that the probability of visiting a doctor due to a common cold is decreased by 0.7 (or 2.1) % points. Both findings can be interpreted as evidence of reduced ex-post moral hazard. Obviously, the amount of the premium refund is high enough to encourage individuals to forgo unnecessary doctor visits.

Effects differ among subgroups. They are mainly driven by individuals aged 49 and under, and men have a stronger reaction than women. By contrast, individuals aged 50 and over do not react to the tariff's incentives at all. A reason why women have a weaker reaction to these incentives might be that they are, in general, more risk-averse than men [37]. Probably, most women prefer a doctor's opinion even in rather harmless situations. Individuals aged 50 and over do, on average, suffer from more severe illnesses compared to younger individuals. For these illnesses, demand is less price-elastic. This explains why they generally do not react to the premium refund's incentives, and is in line with the previous research. Schmitz [2] finds that individuals that had high medical demands in the past—presumably ill individuals—do not react to the expansion of insurance coverage. Likewise, Gerfin et al. [16] observe that healthy individuals react much more strongly to incentives.

Even though I use lagged outcomes as proxy variables for unobserved heterogeneity, one possible weakness of this study is that relevant characteristics cannot be explicitly controlled for (e.g., lifestyle factors). Another limitation is that the data only comprise individuals from two sickness funds which may not be completely representative of all sickness funds in Germany.

This study focuses on contemporaneous effects. Further research is needed to truly identify causal effects, to consider more strongly the nonlinear nature of this scheme, and to find out about the long-term consequences of the tariff's incentives.

Compliance with ethical standards

Conflict of interest I am an employee of Team Gesundheit GmbH. Apart from that, there are no financial, ethical, or other potential conflicts of interests.

Human and animal rights statement This article is based on previously available data, and does not involve any new studies of human subjects performed by the author. However, approvals from the sickness funds were obtained to use their data for this study.

Appendix

See Fig. 2, Tables 9, 10, and 11.

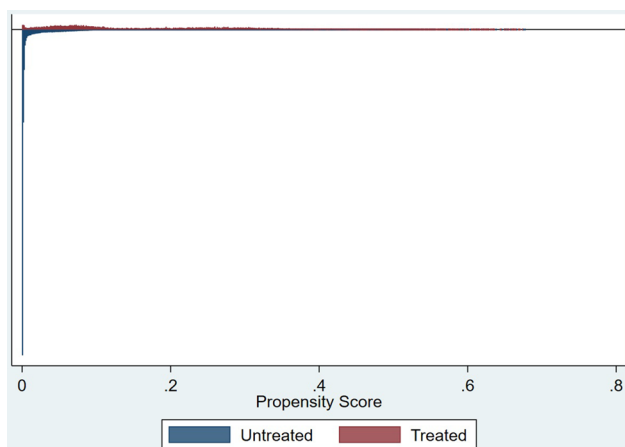[21] The oldest participant is 71 years old.

Fig. 2 Untrimmed propensity score. The figure shows the distribution of the untrimmed propensity score for the treatment and control group

Table 9 Comparison of the sample and the SHI population in 2010

Age groups	Sample: share in %		SHI: share in %	
	Men	Women	Men	Women
14 and younger	0.22	0.18	0.13	0.12
15–19	1.89	1.27	0.78	0.57
20–24	4.15	3.15	2.85	2.62
25–29	4.42	3.35	3.88	3.65
30–34	3.92	3.01	3.78	3.36
35–39	4.06	2.95	3.69	3.18
40–44	5.44	4.05	4.92	4.27
45–49	5.79	4.23	5.39	4.73
50–54	4.89	3.62	4.69	4.19
55–59	3.99	2.91	4.01	3.61
60–64	3.32	2.60	3.42	3.22
65–69	3.32	2.91	3.37	3.80
70–74	3.92	3.54	3.85	4.54
75–79	2.83	2.76	2.47	3.20
80 and older	2.89	4.50	2.43	5.29
Total	54.97	45.03	49.65	50.35

Source: GBE, 2015 and own calculations

Table 10 Linear coefficients of propensity score estimation in main specification

Covariate	
Age	0.096*** (0.006)
Male	0.679*** (0.062)
Bonus program in 2006–2010	0.819*** (0.130)
Sickness fund 2	−1.504*** (0.016)
Compulsorily insured	0.719*** (0.166)
Voluntarily insured	1.114*** (0.171)
Unemployed	0.802 (0.581)
Pensioner	−0.484*** (0.151)
Educational category 2	0.388*** (0.060)
Educational category 3	−0.600*** (0.220)
Educational category 4	0.544*** (0.123)
Educational category 5	0.460*** (0.074)
Educational category 6	0.117*** (0.031)
Positive expenditures total, before	0.403 (0.256)
Expenditures total, before	−0.000 (0.000)
Doctor visit, before	1.593*** (0.169)
Expenditures on doctor visits, before	−0.001*** (0.000)
No. of doctor visits, before	−0.006*** (0.002)
Hospital visit, before	−0.441 (0.360)
Expenditures on hospital visits, before	0.000 (0.000)
Drug prescription, before	0.046 (0.246)
No. of drug prescriptions, before	0.051** (0.025)
Sickness absence, before	0.323*** (0.122)
Expenditures on sickness benefit, before	−0.000 (0.000)
No. of times on sickness absence, before	0.013 (0.115)
No. of days on sickness absence, before	−0.005** (0.002)
Other costs, before	0.000 (0.000)
Common cold (J00), before	0.016 (0.043)
Common cold (J00–J06), before	−0.076*** (0.013)
N treated	13,157
N controls	409,750

Robust standard errors in parentheses. ***$p<0.01$, **$p<0.05$, *$p<0.10$. The table shows the coefficients in the propensity score matching for all linear matching variables in the main specification. Beyond that, a long list of quadratic and interaction terms was used for the matching as well. These coefficients are available upon request from the author. Covariates concerning diagnoses, costs, or other measures of medical utilization are measured during the years 2006–2008. Socioeconomic covariates are measured in 2010

Table 11 Detailed results for the main specification

Coefficient	(1) Doctor visit	(2) Doctor visit (GP)	(3) Doctor visit (specialist)	(4) No. of doctor visits	(5) No. of doctor visits (GP)	(6) No. of doctor visits (specialist)
Premium refund tariff	−0.018*** (0.004)	−0.026*** (0.005)	−0.011* (0.006)	−0.479*** (0.121)	−0.284*** (0.051)	−0.181** (0.091)
Age	0.005* (0.003)	0.003 (0.003)	0.009*** (0.003)	0.002 (0.063)	−0.009 (0.026)	0.032 (0.046)
Male	−0.136*** (0.038)	−0.022 (0.036)	−0.205*** (0.038)	−2.236*** (0.717)	−0.104 (0.184)	−2.058*** (0.641)
Bonus program in 2006–2010	0.081 (0.091)	0.072 (0.094)	0.097 (0.089)	−0.381 (2.033)	0.936 (0.817)	−0.978 (1.617)
Sickness fund 2	−0.041*** (0.010)	−0.054*** (0.012)	−0.042*** (0.012)	−0.299 (0.213)	−0.056 (0.092)	−0.153 (0.163)
Compulsorily insured	0.506*** (0.074)	0.449*** (0.072)	0.375*** (0.080)	3.431** (1.654)	1.415** (0.655)	1.998* (1.178)
Voluntarily insured	0.220** (0.088)	0.176** (0.088)	0.153 (0.096)	0.437 (2.078)	0.437 (0.803)	−0.049 (1.564)
Unemployed	0.439** (0.185)	0.161 (0.233)	0.346* (0.201)	2.270 (2.740)	−0.632 (2.110)	2.618* (1.573)
Pensioner	0.013 (0.103)	0.008 (0.104)	−0.056 (0.100)	0.827 (1.757)	−0.735 (0.589)	1.516 (1.257)
Educational category 2	−0.103 (0.028)	−0.022 (0.034)	−0.077** (0.035)	−1.434* (0.792)	−0.645** (0.303)	−0.789 (0.661)
Educational category 3	0.051 (0.118)	−0.013 (0.112)	0.025 (0.130)	1.300 (3.442)	−1.175* (0.621)	2.285 (2.929)
Educational category 4	0.051 (0.064)	0.002 (0.073)	0.047 (0.069)	−0.283 (1.389)	0.000 (0.466)	−0.359 (1.187)
Educational category 5	−0.025 (0.054)	0.008 (0.055)	0.007 (0.053)	0.031 (0.701)	−0.132 (0.273)	0.186 (0.553)
Educational category 6	−0.025 (0.018)	−0.063*** (0.021)	−0.007 (0.021)	−0.774 (0.762)	−0.591*** (0.156)	−0.216 (0.763)
Positive expenditures total, before	−0.304** (0.142)	−0.390*** (0.120)	−0.202 (0.134)	−0.586 (1.724)	−1.223* (0.698)	0.885 (1.015)
Expenditures total, before	0.000** (0.000)	0.000** (0.000)	0.000 (0.000)	0.001 (0.001)	0.000 (0.000)	0.000 (0.001)
Doctor visit, before	−0.060 (0.117)	−0.095 (0.117)	−0.047 (0.115)	0.527 (1.717)	0.059 (0.750)	0.648 (1.215)
Expenditures on doctor visits, before	0.000*** (0.000)	0.000*** (0.000)	0.000** (0.000)	0.001 (0.005)	−0.001 (0.001)	0.002 (0.004)
No. of doctor visits, before	0.003*** (0.001)	0.005*** (0.001)	0.008*** (0.001)	0.422*** (0.033)	0.153*** (0.012)	0.259*** (0.030)
Hospital visit, before	−0.000 (0.106)	0.107 (0.100)	−0.076 (0.099)	−1.725 (1.544)	−0.881 (0.831)	−0.612 (0.832)
Expenditures on hospital visits, before	0.000 (0.000)	−0.000 (0.000)	−0.000* (0.000)	−0.002** (0.001)	−0.000 (0.000)	−0.002** (0.001)
Drug prescription, before	0.390** (0.127)	0.418*** (0.110)	0.291** (0.127)	1.165 (1.667)	0.697 (0.721)	0.001 (0.884)
No. of drug prescriptions, before	0.021 (0.024)	0.024 (0.023)	0.048* (0.025)	−0.128 (0.352)	−0.222* (0.118)	0.095 (0.282)
Sickness absence, before	−0.139** (0.070)	−0.077 (0.068)	−0.103 (0.072)	−1.368 (1.177)	−0.100 (0.712)	−1.264* (0.685)
Expenditures on sickness benefit, before	0.000** (0.000)	0.000 (0.000)	0.000** (0.000)	−0.003*** (0.001)	−0.001*** (0.000)	−0.002** (0.001)
No. of times on sickness absence, before	−0.136*** (0.037)	−0.108*** (0.034)	−0.082** (0.037)	−1.087* (0.630)	−0.855** (0.401)	−0.272 (0.310)
No. of days on sickness absence, before	−0.002** (0.001)	−0.002* (0.001)	−0.001 (0.001)	−0.059** (0.028)	−0.017 (0.012)	−0.041* (0.022)
Other costs, before	−0.000** (0.000)	−0.000* (0.000)	−0.000** (0.000)	−0.002 (0.001)	−0.001 (0.001)	−0.001 (0.001)
Common cold (J00), before	0.027 (0.017)	0.045** (0.020)	0.046* (0.024)	−0.410 (0.445)	−0.327* (0.194)	−0.071 (0.325)
Common cold (J00–J06), before	0.027*** (0.006)	0.042*** (0.007)	0.035*** (0.008)	0.417*** (0.147)	0.082 (0.063)	0.320*** (0.111)
N treated	5090					
N controls	360,251					
N controls (weighted)	5069					

Table 11 (continued)

Coefficient	(7) Common cold (J00)	(8) Common cold (J00–J06)	(9) Expenditures on doctor visits	(10) Expenditures on doctor visits (GP)	(11) Expenditures on doctor visits (specialist)
Premium refund tariff	−0.007*** (0.002)	−0.021*** (0.006)	−13.227*** (4.443)	−6.883*** (1.236)	−5.060 (3.920)
Age	−0.001 (0.001)	0.001 (0.003)	2.443 (2.474)	−0.875 (0.661)	3.530 (2.204)
Male	−0.002 (0.004)	−0.061** (0.027)	−87.778** (36.688)	−2.955 (5.436)	−84.763** (35.860)
Bonus program in 2006–2010	−0.004 (0.019)	−0.124 (0.080)	60.404 (75.907)	23.294 (21.555)	50.642 (69.783)
Sickness fund 2	0.002 (0.003)	−0.022** (0.010)	−10.413 (8.112)	−1.344 (2.414)	−8.411 (7.150)
Compulsorily insured	0.020 (0.017)	0.167** (0.067)	127.631** (63.574)	45.791*** (15.457)	83.107 (56.683)
Voluntarily insured	0.006 (0.021)	−0.071 (0.083)	31.150 (79.622)	29.337 (20.141)	−2.643 (71.955)
Unemployed	0.005 (0.023)	0.177 (0.150)	200.895** (98.282)	11.209 (42.168)	187.600* (104.063)
Pensioner	−0.000 (0.012)	0.003 (0.031)	12.459 (61.658)	−22.715 (20.333)	34.617 (45.507)
Educational category 2	−0.003 (0.009)	0.016 (0.037)	−41.735 (29.777)	−10.067 (7.386)	−32.055 (27.761)
Educational category 3	−0.024 (0.026)	0.092 (0.129)	88.857 (145.950)	−16.943 (16.588)	107.143 (140.128)
Educational category 4	−0.000 (0.019)	0.087 (0.066)	−45.562 (61.758)	4.350 (12.211)	−47.671 (58.702)
Educational category 5	−0.007 (0.005)	−0.042 (0.032)	−3.451 (25.118)	−0.237 (7.462)	1.546 (22.417)
Educational category 6	−0.000 (0.004)	−0.006 (0.019)	−42.699 (41.939)	−13.683*** (3.885)	−25.674 (42.498)
Positive expenditures total, before	0.004 (0.007)	0.045 (0.080)	−26.399 (52.936)	−44.655** (18.703)	25.480 (39.944)
Expenditures total, before	0.000* (0.000)	0.000 (0.000)	−0.008 (0.028)	0.007 (0.009)	−0.014 (0.026)
Doctor visit, before	−0.012 (0.011)	−0.065 (0.079)	60.901 (73.308)	−3.581 (23.102)	64.405 (62.979)
Expenditures on doctor visits, before	0.000 (0.000)	0.000* (0.000)	0.153 (0.250)	0.018 (0.037)	0.133 (0.256)
No. of doctor visits, before	−0.000 (0.000)	0.001 (0.001)	4.542*** (1.514)	2.196*** (0.330)	2.346 (1.569)
Hospital visit, before	−0.017** (0.007)	−0.076 (0.080)	−41.845 (47.883)	−14.292 (23.395)	−21.344 (35.876)
Expenditures on hospital visits, before	−0.000 (0.000)	−0.000 (0.000)	−0.040 (0.033)	0.004 (0.013)	−0.045 (0.028)
Drug prescription, before	−0.011* (0.006)	−0.058 (0.071)	33.998 (47.637)	28.700 (19.572)	−3.482 (32.435)
No. of drug prescriptions, before	0.003 (0.004)	0.037** (0.019)	−8.459 (13.420)	−2.727 (3.126)	−5.218 (12.469)
Sickness absence, before	−0.002 (0.011)	−0.051 (0.050)	−21.293 (43.760)	6.357 (23.591)	−26.512 (31.038)
Expenditures on sickness benefit, before	−0.000 (0.000)	0.000 (0.000)	−0.065 (0.041)	−0.025* (0.014)	−0.044 (0.035)
No. of times on sickness absence, before	−0.007* (0.004)	−0.048* (0.025)	−36.302* (21.771)	−24.238* (13.731)	−13.801 (12.039)
No. of days on sickness absence, before	0.000 (0.000)	−0.004*** (0.001)	−1.855 (1.244)	−0.590* (0.327)	−1.344 (1.146)
Other costs, before	−0.000 (0.000)	−0.000 (0.000)	−0.065 (0.053)	−0.043** (0.019)	−0.032 (0.043)
Common cold (J00), before	0.051*** (0.012)	0.021 (0.027)	−15.176 (17.233)	−2.349 (5.152)	−11.190 (14.642)
Common cold (J00–J06), before	0.008*** (0.003)	0.119*** (0.008)	18.886*** (5.584)	2.177 (1.538)	16.571*** (5.008)
N treated	5090				
N controls	360,251				
N controls (weighted)	5069				

Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.10. Each column contains the coefficients of one regression. To all regressions applies: bandwidth=0.005, trimming at 0.1. The table shows the coefficients in the regression adjustment for all linear covariates in the main specification. Beyond that, a long list of quadratic and interaction terms was used in the regressions as well. These coefficients are available upon request from the author. Covariates concerning diagnoses, costs, or other measures of medical utilization are measured during the years 2006–2008. Socioeconomic covariates and all outcomes are measured in 2010

GP general practitioner, *No.* number

References

1. OECD: Focus on Health Spending. OECD Health Statistics 2015. https://www.oecd.org/health/health-systems/Focus-Health-Spending-2015.pdf (2015). Accessed 20 March 2016

2. Schmitz, H.: More health care utilization with more insurance coverage? Evidence from a latent class model with German data. Appl. Econ. **44**, 4455–4468 (2012)

3. Arrow, K.J.: Uncertainty and the welfare economics of medical care. Am. Econ. Rev. **53**, 941–973 (1963)

4. Pauly, M.: The economics of moral hazard: comment. Am. Econ. Rev. **58**, 531–537 (1968)

5. Manning, W.G., Newhouse, J.P., Duan, N., Keeler, E.B., Leibowitz, A.: Health insurance and the demand for health care. Evidence from a randomized experiment. Am. Econ. Rev. **77**, 251–277 (1987)

6. Health Policy Brief: The Oregon health insurance experiment, Health Affairs (2015). https://doi.org/10.1377/hpb2015071 6.236899

7. Chandra, A., Gruber, J., McKnight, R.: Patient cost-sharing and hospitalization offsets in the elderly. Am. Econ. Rev. **100**, 192–213 (2010)

8. Boes, S., Gerfin, M.: Does full insurance increase the demand for health care? Health Econ. **25**, 1483–1496 (2016)

9. Farbmacher, H., Winter, J.: Per-period co-payments and the demand for health care: evidence from survey and claims data. Health Econ. **22**, 1111–1123 (2013)

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

11. Kunz, J.S., Winkelmann, R.: An econometric model of healthcare demand with non-linear pricing. Health Econ. **26**, 691–702 (2017)

12. Felder, S., Werblow, A.: A physician fee that applies to acute but not to preventive care: evidence from a German deductible program. Schmollers Jahrbuch **128**, 191–212 (2008)

13. Hemken, N., Schusterschitz, C., Thöni, M.: Optional deductibles in GKV (statutory German health insurance): do they also exert an effect in the medium term? J. Public Health **20**, 219–226 (2012)

14. Pütz, C., Hagist, C.: Optional deductibles in social health insurance systems: findings from Germany. Eur. J. Health Econ. **7**, 225–230 (2006)

15. Hayen, A., Klein, T.J., Salm, M.: Does the framing of patient cost-sharing incentives matter? The effects of deductibles vs. no-claim refunds. IZA DP 11508 (2018)

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

17. Aron-Dine, A., Einav, L., Finkelstein, A., Cullen, M.: Moral hazard in health insurance: do dynamic incentives matter? Rev. Econ. Stat. **97**, 725–741 (2015)

18. Altonji, J.G., Elder, T.E., Taber, C.R.: Selection on observed and unobserved variables: assessing the effectiveness of catholic schools. J. Polit. Econ. **113**, 151–184 (2005)

19. Oster, E.: Unobservable selection and coefficient stability: theory and evidence. Unpublished working paper. https://www.brown.edu/research/projects/oster/sites/brown.edu.research.projects.oster/files/uploads/Unobservable_Selection_and_Coefficient_Stability_0.pdf (2016). Accessed 21 Feb. 2019

20. Koç, Ç.: Disease-specific moral hazard and optimal health insurance design for physician services. J. Risk Insur. **78**, 413–446 (2011)

21. GBE, Gesundheitsberichterstattung des Bundes. https://www.gbe-bund.de (2015). Accessed 13 August 2015

22. Federal Statistical Office. https://www.destatis.de (2016). Accessed 6 February 2016

23. GKV head organization. https://www.gkv-spitzenverband.de (2016). Accessed 6 February 2016

24. Malin, E.M., Schmidt, E.M.: Beitragsrückzahlung in der GKV: ein Instrument zur Kostendämpfung? Die Betriebskrankenkasse **12**(1995), 759–763 (1995)

25. Cutler, D.M., Finkelstein, A., McGarry, K.: Preference heterogeneity and insurance markets: explaining a puzzle of insurance. Am. Econ. Rev. Papers Proc. **98**, 157–162 (2008)

26. Wooldridge, J.M.: Introductory econometrics: a modern approach, 6th edn. South Western Educational Publ, Mason (2015)

27. García Gómez, P., López Nicolás, A.: Health shocks, employment and income in the Spanish labour market. Health Econ. **15**, 997–1009 (2006)

28. García-Gómez, P.: Institutions, health shocks and labour market outcomes across Europe. J. Health Econ. **30**, 200–213 (2011)

29. Hagan, R., Jones, A.M., Rice, N.: Health and retirement in Europe. Int. J. Environ. Res. Public Health **6**, 2676–2695 (2009)

30. Imbens, G.W.: Matching methods in practice: three examples. J. Hum. Resour. **50**, 373–419 (2015)

31. Einav, L., Finkelstein, A., Ryan, S.P., Schrimpf, P., Cullen, M.R.: Selection on moral hazard in health insurance. Am. Econ. Rev. **103**, 178–219 (2013)

32. Finkelstein, A., Arrow, K.J., Gruber, J., Newhouse, J.P., Stiglitz, J.E.: Moral hazard in health insurance, 1st edn. Columbia University Press, New York (2015)

33. Imbens, G.W., Wooldridge, J.M.: Recent developments in the econometrics of program evaluation. J. Econ. Lit. **47**, 5–86 (2009)

34. Schmitz, H., Westphal, M.: Short- and medium-term effects of informal care provision on female caregivers' health. J. Health Econ. **42**, 174–185 (2015)

35. Marcus, J.: Does job loss make you smoke and gain weight? Economica **81**, 626–648 (2014)

36. Caliendo, M., Kopeinig, S.: Some practical guidance for the implementation of propensity score matching. J. Econ. Surv. **22**, 31–72 (2008)

37. Borghans, L., Golsteyn, B.H.H., Heckman, J.J., Meijers, H.: Gender differences in risk aversion and ambiguity aversion. J. Eur. Econ. Assoc. **7**, 649–658 (2009)