HEDG Working Paper 09/17

The effect of private health insurance on medical care utilization and self-assessed health in Germany

Patrick Hullegie

Tobias J. Klein

July 2009

ISSN 1751-1976

york.ac.uk/res/herc/hedgwp

The effect of private health insurance on medical care utilization and self-assessed health in Germany*

Patrick Hullegie† and Tobias J. Klein‡

July 2009

Abstract

In Germany, employees are generally obliged to participate in the public health insurance system, where coverage is universal, co-payments and deductables are moderate, and premia are based on income. However, they may buy private insurance instead if their income exceeds the compulsory insurance threshold. Here, premia are based on age and health, individuals may choose to what extent they are covered, and deductables and co-payments are common. In this paper we estimate the effect of private insurance coverage on the number of doctor visits and self-assessed health. Variation in income around the compulsory insurance threshold provides a natural experiment that we exploit to control for selection into private insurance. We document that income is measured with error and suggest an approach to take this into account. We find negative effects of private insurance coverage on the number of doctor visits and positive effects on health.

JEL Classification: I11, I12, C31.
Keywords: Private health insurance, medical care utilization, selection into insurance, natural experiment, regression discontinuity design, measurement error.

*We would like to thank Jaap Abbring, Katie Carman, Peter Kooreman, Martin Salm and participants of the Netspar seminar for helpful comments.

†Netspar, CentER, Tilburg University. *Address:* Tilburg University, Department of Economics, PO Box 90153, 5000 LE Tilburg, The Netherlands. *E-Mail:* P.G.J.Hullegie@uvt.nl.

‡Netspar, CentER, Tilburg University. *Address:* Tilburg University, Department of Econometrics and OR, PO Box 90153, 5000 LE Tilburg, The Netherlands. *E-Mail:* T.J.Klein@uvt.nl.

1 Introduction

In Germany, employees are generally obliged to participate in the public health insurance system, where coverage is universal, co-payments and deductables are moderate, and premia are based on income. However, they may buy private insurance instead if their income exceeds the so-called compulsory insurance threshold.[1] Here, premia are based on age and health, individuals may choose to what extent they are covered, and deductables and co-payments are common. These differences in the incentive structure may affect the demand for medical care even if the treatment provided to privately and publicly insured patients is exactly the same. In that situation we would expect privately insured patients to be less inclined to demand medical services.

An important difference affecting the supply of services is that for the same treatment the compensation doctors receive for privately insured patients is 2.3 times as high as the compensation for publicly insured patients. For that reason doctors have an incentive to treat privately insured patients first, and more intensely, possibly providing better treatment (Jürges, 2009). For example, waiting times for privately insured patients are lower on average (Lungen et al., 2008). This may in turn affect the demand for medical care. Ultimately, it is an empirical question whether the amount of services provided is higher or lower for privately insured individuals, and whether insurance type has an effect on health.

In this paper, we study the effect of being privately insured on the number of doctor visits and self-assessed health. We control for selection into private insurance by exploiting an unusual feature of the German public health insurance system. As soon as income in the last year exceeds the so-called compulsory insurance threshold individuals become eligible to opt out of the public health insurance system and may buy private insurance. Random variation in income around this compulsory insurance threshold generates a natural experiment, which we exploit to estimate the effect of private health insurance on two measures of medical care utilization, namely the

[1] About 90 per cent of the German population is insured in the public health insurance system. Most remaining individuals buy private insurance (Colombo & Tapay, 2004).

number of doctor visits and the nights spent in a hospital, and on self-assessed health. This effect is estimated for the subgroup of individuals who opt out of the public system and buy private insurance.

We use survey data from the German Socio Economic Panel (GSOEP) for our analysis because German administrative data, that contain accurate income measures, do not contain health related information. In the data we find that there is a sizable number of individuals who, according to their reported income, are not eligible to buy private insurance but at the same time report to be privately insured. This suggests that income, which determines eligibility, is measured with error. The methodological contribution in this paper is to model the measurement error, which then allows us to estimate the effect of interest. For this we assume that the measurement error is normally distributed and classical, and exploit the fact that the probability to buy private insurance is zero for ineligibles.

Controlling for selection into private insurance we find a significant negative effects of being privately insured on the number of doctor visits. At the same time we find that being privately insured has a significant positive effect on self-assessed health. An explanation for this could be that doctors allocate more time to treating privately insured patients, which results in better treatment.[2]

The remainder of this paper is organized as follows. Section 2 discusses related results. The institutional details are discussed in Section 3. Section 4 discusses the econometric approach, emphasizing our approach to modeling measurement error. Section 5 contains a description of the data. Results are presented in Section 6, and a sensitivity analysis is performed in Section 7. Finally, Section 8 concludes.

[2]As already pointed out before they have strong financial incentives to do so. Furthermore, there is indirect evidence that doctors face strong time constraints when treating patients. The consultation length for the average (publicly insured) individual is very low in Germany. Deveugele, Derese, van den Brink-Muinen, Bensing, and De Maeseneer (2002, Table 4) compare the average consultation length for general practitioners in six countries and find that with 7.6 minutes it is lowest in Germany. It is highest in Switzerland, where it is equal to 15.6 minutes.

2 Related Literature

The empirical literature on demand for health services dates at least back to the 1970s, when the RAND Health Insurance Experiment (HIE) was conducted. Findings of this randomized experiment are that the use of medical services responds to changes in cost sharing, with a stronger effect for outpatient care than for inpatient care (Newhouse, 1974; Manning et al., 1987). Additionally, insurance coverage is found to have an insignificant impact on health status of the overall population (Newhouse, 1993). Conducting such a randomized experiment is, however, typically not feasible because of financial constraints, ethical considerations, or other reasons. The literature on health care utilization that uses observational data is vast and summarized by Cutler and Zeckhauser (2000), Levy and Meltzer (2004), and Buchmüller et al. (2005).

There are at least three studies for Germany that relate demand for medical services to insurance type. Geil et al. (1997) estimate a count data model for hospital visits and find no relationship between insurance coverage and the hospitalization decision. Riphahn et al. (2003) estimate a bivariate count data model for physician and hospital visits and find that neither hospital nights nor doctor visits depend on the insurance type of the individual. Finally Jürges (2009) finds that privately insured are less likely to contact a doctor, but conditional on seeing a doctor at least once the number of doctor visits is significantly larger than that of patients covered by public health insurance. These studies, however, suffer from the problem that they cannot fully control for selection into private insurance. In general, disentangling selection into insurance and the effect of a particular insurance on behavior is difficult.[3]

An approach that has become popular to control for arbitrary selection into health insurance is to exploit *natural experiments*, a setting in which a certain group of individuals experi-

[3]Until recently both the theoretical and the empirical literature on informational asymmetries focused on *adverse* selection and moral hazard (Akerlof, 1970; Rothschild & Stiglitz, 1976; Arrow, 1963). However, Finkelstein and McGarry (2006) and Fang et al. (2008) point out that there might be *advantageous* selection instead. Their explanation is that good risks select into insurance because they are more risk averse and therefore value insurance more than bad risks do.

ences an unexpected and exogenous change in the incentive structure, and to perform a *regression discontinuity analysis*. The regression discontinuity (RD) approach has been suggested by Thistlethwaite and Campbell (1960) and has recently been developed by Hahn et al. (2001).[4] They show that under relatively mild assumptions the RD method can be interpreted as a local randomized experiment. The advantage is that the results have a strong internal validity, but the drawback is that the study population if often only a small subset of the population of interest (or the population that a social planner must treat). In the context of health insurance Chiappori et al. (1998) analyze data from a French natural experiment and find no evidence of moral hazard for general practitioner or specialists office visits. Two recent applications of the RD approach to health insurance are Card et al. (2008) and Card et al. (2009), who exploit the fact that in the United States individuals of age 65 and above are generally insured because they are eligible for Medicare. Finding are that this leads to an increase in doctor visits with the highest increase for groups that previously lacked coverage. There is a sharp increase in hospitalization but effects differ across groups and by type of admission. They find an increase in the number of procedures performed, treatment intensity, as well as a significant and large reduction in mortality.

3 Institutional details

In Germany, about 90% of the population is publicly insured (Colombo & Tapay, 2004). Buying public insurance is mandatory for dependent employees as long as their income does not exceed the so-called compulsory insurance threshold. The public insurance premium equals a certain percentage (nowadays about 14 percent that are shared between the employer and the employee) of gross income up to the so-called contribution ceiling, and equal to it thereafter.[5]

Table 1 shows the contribution ceilings and the compulsory insurance thresholds by the year

[4] See also Imbens and Lemieux (2008), Lee and Lemieux (2009) and Van der Klaauw (2009) for recent discussions.

[5] See Jürges (2009) and the references therein for more details.

Year	Contribution ceiling	Compulsory insurance threshold
2000	39,576	39,576
2001	40,032	40,032
2002	40,500	40,500
2003	41,400	45,900
2004	41,850	46,350
2005	42,300	46,800
2006	42,750	47,250

Reported for West-Germany. Amounts are per year and in Euros. Reported by year in which the income was earned.

Table 1: Contribution ceiling and compulsory insurance threshold

in which the income was earned. To give an example of how the system works consider an individual whose income, including all extra payments, in 2000 was $40,000$ Euros. Then he is eligible to buy private insurance in 2001 because his income exceeded $39,576$ Euros, the compulsory insurance threshold. The insurance premium in the public system is based on the contribution ceiling, which was equal to the compulsory insurance threshold until 2002. If his income stays the same in 2001, then he will have to join the public insurance again in 2002 because the compulsory insurance threshold is $40,032$ Euros for income earned in 2001.

Due to a reform the compulsory insurance threshold increased substantially for income earned in 2003 and later. The contribution ceiling increased only moderately. To avoid that a substantial number of individuals needed to switch back to buying public insurance in 2004 because of this increase in the compulsory insurance threshold for income earned in 2003 those individuals who actually bought private insurance in 2002 but, according to the new thresholds, are not eligible for this any more could still do so provided that their income is at least equal to the contribution ceiling.[6]

Contributions for private health insurance are mainly based on health and age. As a conse-

[6]In the empirical analysis we incorporate this by setting their income to the contribution ceiling.

quence of this, and because of the fact that private insurers are allowed to reject individuals, the risk pool of the private insurers is much better than in the public system and therefore contributions are on average lower.

In the public system coverage is universal. Deductables and co-payments are limited. Privately insured individuals can buy better care, e.g. treatment by the head doctor in a hospital or a single room in a hospital, but this comes at a higher price. Deductibles and co-payments are much more common, and many insurers offer a rebate if an individual does not use medical services in the past calendar year. Unfortunately, specific characteristics of private insurance are unobserved in our data.

4 Econometric approach

4.1 Regression discontinuity design

Let $(y_i(0), y_i(1))$ be the pair of potential outcomes for each member i of the study population. In our case $y_i(0)$ denotes the health outcome one would experience in case of public health insurance and $y_i(1)$ denotes the health outcome one would experience with private health insurance. That is we consider private health insurance to be the "treatment."

An individual is eligible to buy private health insurance instead of public insurance if his income in the previous year exceeded the respective compulsory insurance threshold. Eligibility thus depends on the deterministic rule $z_i^* \geq 0$, where z_i^* denotes the difference between income earned in the previous year and the compulsory insurance threshold for the current year. For $z_i^* < 0$ individuals in our sample are not eligible to buy private insurance, whereas for $z_i^* \geq 0$ they are. Buying private insurance is voluntary for eligible individuals so that some will buy it while others will not and consequently we can divide the study population into three subgroups: *ineligibles*, *eligible non-participants*, and *eligible participants*.

Hahn et al. (2001) show that two assumptions are needed to identify a local average treatment effect. First, the mean value of $y_i(0)$ conditional on z_i^* is a continuous function of z_i^* at $z_i^* = 0$. Second, we need that the decision to buy private insurance is monotone in eligibility. This is the monotonicity condition of Imbens and Angrist (1994). It holds by construction because ineligibles cannot buy private insurance. Under these assumptions

$$\Delta^{LATE} \equiv \mathbb{E}\left(y_i(1) - y_i(0) \,|\, p_i = 1, z_i^* = 0\right) = \frac{\mathbb{E}\left(y_i | z_i^* = 0^+\right) - \mathbb{E}\left(y_i | z_i^* = 0^-\right)}{\mathbb{E}\left(p_i | z_i^* = 0^+\right)}, \tag{1}$$

where y_i is the *observed* health outcome, p_i is an indicator of private insurance, $\mathbb{E}\left(\cdot | z_i^* = 0^+\right) \equiv \lim_{\delta \downarrow 0} \mathbb{E}\left(\cdot | z_i^* = \delta\right)$, and $\mathbb{E}\left(\cdot | z_i^* = 0^-\right) \equiv \lim_{\delta \uparrow 0} \mathbb{E}\left(\cdot | z_i^* = \delta\right)$. See also Battistin and Rettore (2008). This is the average treatment effect for those individuals who buy private insurance when becoming eligible.

4.2 Measurement error

The data contains a sizable number of individuals who according to their reported income are not eligible to buy private insurance, but at the same time report to have done so.[7] Misreporting insurance status or measurement error in income may both be valid explanations for this.[8]

We consider it to be more plausible that income is measured with error because income is a real number, and may thus be recalled with errors, whereas insurance status is more easily known because it is typically either public or private insurance. Moreover, there is direct evidence for measurement error in income because the GSOEP questionnaire asks respondents twice about their monthly income in a given year.[9] In a given year respondents are asked about their current

[7] Our data contains 75 ineligible individuals with private insurance, which amounts to 15% of the individuals with private insurance. Eligibility status is computed on the basis of self-reported income.

[8] There is an extensive literature on measurement error in income, see for example Bound et al. (2001) for a survey. In order to study the accuracy of survey reports, they are typically compared with either employers' or administrative records. Some studies find that survey reports are highly correlated with record values, while others find much lower correlations. The mean of survey reports is found to be close to the mean of the record values. Under- or over-reporting, if present, is found to be moderate on average. Kreider (forthcoming) and Kreider and Hill (forthcoming) are two recent papers studying classification error in health insurance status.

[9] This is not the case for the total yearly income that we use to determine eligibility. Yearly income includes extra

monthly income and about their income in the previous year. The question in the former case is:

> *"How high was your income from employment last month? If you received extra income such as vacation pay or back pay, please do* **not** *include this. Please do include overtime pay."*

The question for the previous calendar year is

> *"We have already asked for your current income. In addition, please state what sources of income you have received in the past calendar year 2005, independent of whether the income was received all year or only in certain months. Look over the list of income sources and check all that apply. For all sources that apply please indicate how many months you received this income in 2005 and how much this was on average per month. Please state the gross amount which means not including deductions or social security."*[10]

This provides us with two measures of monthly income for the same year. If both income reports would be reported without any error, and if the within year variance in monthly income is low, then both measures should be close to one another, i.e. the data points in a scatter plot should be close to the 45 degree line. Such a scatter plot is shown in Figure 1. The deviations from the 45 degree line are substantial. Figure 2 are histograms for either of the two measures, holding the other measure fixed at $3,000$ Euros. We interpret this as evidence for measurement error in income.

Measurement error in income leads to misclassification of eligibility. Importantly, this misclassification is not independent of the true underlying income because if the true underlying income is below (above) the compulsory insurance threshold the classification error can only be that the individual is (not) eligible to buy private insurance. This precludes the use of an

payments such as holiday pay.

[10]Both question are taken from the English translation of the 2006 GSOEP personal questionnaire. Available at http://panel.gsoep.de/soepinfo2007/.

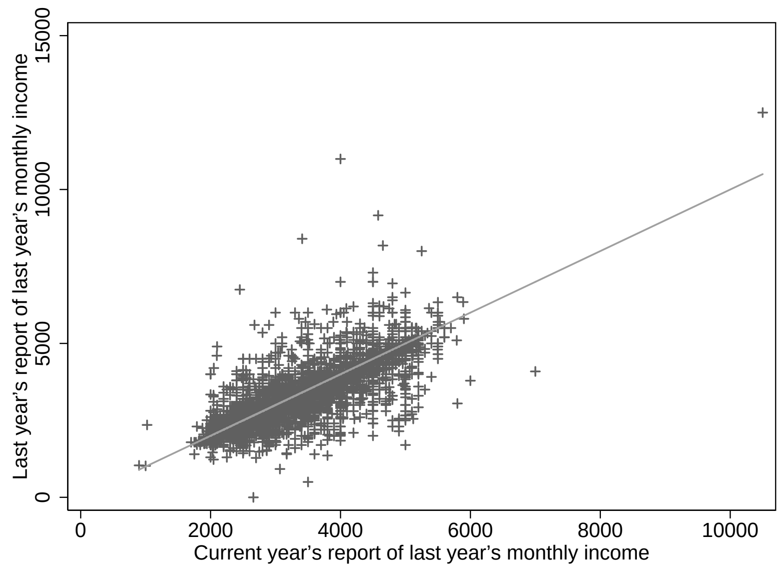

Figure 1: Scatter plot of two income measures

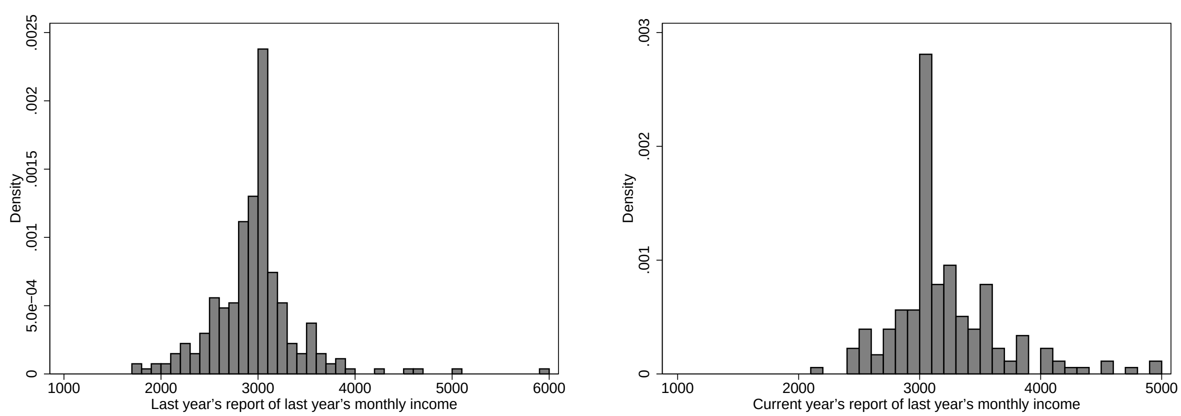

Figure 2: Histograms of monthly income

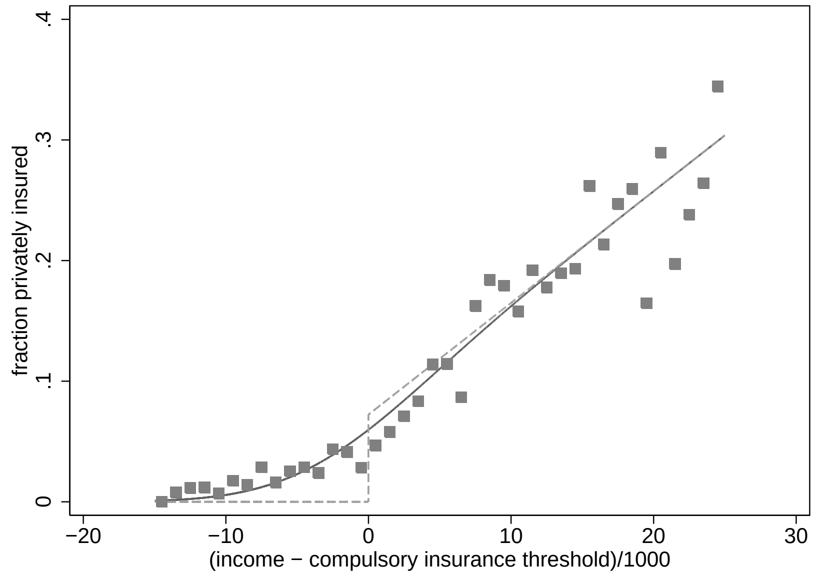

Figure 3: Fraction privately insured, pooled across all years

instrumental variables approach to estimating the unknown quantities in the numerator and denominator in equation (1), respectively.

The effect of measurement error on estimates of these quantities will be that no discontinuity will be observed at the threshold. In Figure 3 the dots are fractions of privately insured individuals which we plot against the difference in income and the compulsory insurance threshold. The figure shows that these fractions are not zero if income is below the compulsory insurance threshold, i.e. if the value of the difference on the horizontal axis is negative, and that indeed there is no discontinuity in the fraction of privately insured at the threshold.[11]

Towards estimating the local average treatment effect in the presence of measurement error

[11] Battistin et al. (forthcoming) show that one can estimate the local average treatment effect using standard techniques if the forcing variable, z_i^*, is observed without error for a part of the population. Then, estimates of both the numerator and denominator in equation equation (1) will be biased, but the bias cancels out. In particular, under conditions given in Battistin et al. (forthcoming),

$$\mathbb{E}\left(y_i(1) - y_i(0)\,|\,p_i = 1, z_i^* = 0^+\right) = \frac{\mathbb{E}(y_i|z_i^* = 0^+) - \mathbb{E}(y_i|z_i^* = 0^-)}{\mathbb{E}(p_i|z_i^* = 0^+) - \mathbb{E}(p_i|z_i^* = 0^-)} = \frac{\mathbb{E}(y_i|z_i = 0^+) - \mathbb{E}(y_i|z_i = 0^-)}{\mathbb{E}(p_i|z_i = 0^+) - \mathbb{E}(p_i|z_i = 0^-)}.$$

we now develop an expression for the probability to be privately insured, which is equal to the conditional expectation of the indicator for being privately insured. Our approach is parametric and our main assumption is that $z_i = z_i^* + u_i$, where u_i is classical measurement error that is normally distributed independent of z_i^* and has a mean of zero and variance σ_u^2. Furthermore, u_i is assumed to be independent of private insurance status and the potential outcomes. To simplify the notation we abstract from covariates. We specify

$$\mathbb{E}\left(p_i|z_i^*\right) = \begin{cases} 0 & \text{if } z_i^* < 0 \\ \alpha + \beta z_i^* & \text{if } z_i^* \geq 0. \end{cases} \tag{2}$$

Recall that when true income is below the compulsory insurance threshold, i.e. $z_i^* < 0$, then the probability of being privately insured is zero because ineligibles may not buy private insurance. Conversely, when true income exceeds the compulsory insurance threshold, i.e. $z_i^* \geq 0$, individuals may buy private insurance. We assume a linear probability model for this. The dashed line in Figure 3 is the estimated underlying relationship, if there was no measurement error, for our data.

We show in Appendix A that under these assumptions

$$\mathbb{E}\left(p_i|z_i\right) = \Phi\left(\frac{z_i}{\sigma_u}\right) \cdot \left(\alpha + \beta z_i + \beta \sigma_u \frac{\phi\left(\frac{z_i}{\sigma_u}\right)}{\Phi\left(\frac{z_i}{\sigma_u}\right)}\right),$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $\phi(\cdot)$ is the standard normal probability density function. Notably, this is the prediction for the observed relationship between insurance status and reported income. The solid line in Figure 3 shows the estimated relationship for our data.

A similar expression can be obtained for $\mathbb{E}\left(y_i|z_i\right)$. This involves specifying different linear functions to the left and right of the discontinuity,[12]

[12]From (2) it follows that this basically assumes that for observations whose z_i is close to 0 we have that

$$\mathbb{E}\left(y_i|z_i^*\right) = \begin{cases} \alpha_0 + \beta_0 z_i^* & \text{if } z_i^* < 0 \\ \alpha_1 + \beta_1 z_i^* & \text{if } z_i^* \geq 0. \end{cases}$$

Under the assumptions given it can be shown in a similar way that this results in,

$$\mathbb{E}\left(y_i|z_i\right) = \left(1 - \Phi\left(\frac{z_i}{\sigma_u}\right)\right)\left(\alpha_0 + \beta_0 z_i - \beta_0 \sigma_u \frac{\phi\left(\frac{z_i}{\sigma_u}\right)}{1 - \Phi\left(\frac{z_i}{\sigma_u}\right)}\right)$$
$$+ \Phi\left(\frac{z_i}{\sigma_u}\right)\left(\alpha_1 + \beta_1 z_i + \beta_1 \sigma_u \frac{\phi\left(\frac{z_i}{\sigma_u}\right)}{\Phi\left(\frac{z_i}{\sigma_u}\right)}\right).$$

The parameters for both $\mathbb{E}\left(p_i|z_i\right)$ and $\mathbb{E}\left(y_i|z_i\right)$ will be jointly estimated using the feasible generalized nonlinear least squares estimator for nonlinear systems of equations. We then calculate the local average treatment effect from these parameter estimates. For this observe that α, α_0, and α_1 are equal to $\mathbb{E}\left(p_i|z_i^* = 0^+\right)$, $\mathbb{E}\left(y_i|z_i^* = 0^-\right)$, and $\mathbb{E}\left(y_i|z_i^* = 0^+\right)$, respectively. Hence, it follows from equation (1) that the local average treatment effect is given by

$$\Delta^{LATE} = \frac{\alpha_1 - \alpha_0}{\alpha}. \tag{3}$$

5 Data

The data we use in this study are taken from the German Socio Economic Panel (GSOEP), which contains information at the individual level on medical care utilization, self-assessed health, health insurance characteristics and background variables. We analyze data for the period from 2000 to 2006. We select our sample such that eligibility to opt out of the public insurance system is exclusively determined by income.

We select individuals from West-Germany only and exclude women because the sample of

$\mathbb{E}(y_i(0)|z_i^*) = \alpha_0 + \beta_0 z_i^*$ and that Δ^{LATE} does not depend on z_i^*.

working women is a selected sample. Unemployed individuals who receive unemployment benefits are required to be in the public health insurance system. For them there is no way to opt out and therefore they are excluded. For self-employed, civil servants, soldiers, teachers in private schools and students it is not mandatory to be in the public system, even if their income is below the compulsory insurance threshold. Hence eligibility does not depend on income and therefore they are excluded from the sample as well. Retired individuals, who receive a public pension, are required to have public health insurance. They may opt out if insurance was not mandatory in at least five years after the age of 55 and most of the time before that. Hence eligibility is only weakly related to income and therefore they are excluded. Individuals of age 55 and older are excluded because for them various ways to opt for (early) retirement exist. By excluding these individuals we concentrate on those for whom retiring is not an option. Individuals under the age of 25 are excluded because they are typically covered by their parent's insurance.

To summarize, our sample consists of West-German men, aged 25 to 55, with a regular employment contract for whom eligibility to opt out of the public health insurance system is exclusively determined by income.

One key variable in our analysis is gross yearly income. This is constructed from the respondents' reports on their average gross monthly income in the previous year and their reports on supplementary income such as 13th month salary, 14th month salary, Christmas bonus, vacation pay, profit share, premia, and bonuses.

Insurance status, p_i, is determined by the following question:

> *"How are you insured for sickness: Do you have state health insurance or are you almost exclusively privately insured?"*

A cross-table of eligibility according to the reported income and health insurance type is given in Table 2. It shows that there are 75 individuals that have private health insurance but are not eligible for that. These are 15% of the individuals with private health insurance. Figure 3 visualize this as well. The dots are fractions of privately insured. To the left of zero these

	Public insurance	Private insurance	Total
Ineligible	3722	75	3797
Eligible	2275	416	2691
Total	5997	491	6488

Table 2: Eligibility and health insurance type

fractions should be zero, at least of both insurance status and income were correctly reported. This illustrates that there is a need to model the measurement error, which we assume to be classical and in income. Otherwise, our estimates run the risk of being biased.

Our main outcomes of interest are doctor visits in the past three months and self-assessed health. The respective questions are[13]

1. *"Have you gone to a doctor within the last three months?"*
2. *"How would you describe your current health?"* Very good, good, satisfactory, poor, bad.

Notice that for the health question, 'Very good' is coded as a 1, 'good' as 2, and so on, up to 5. Hence, a positive association between health and private insurance would be reflected in a negative coefficient on an indicator for private insurance in an ordinary least squares regression.

In a sensitivity analysis we also look at the number of nights spent in a hospital. The corresponding question is

"How many nights altogether did you spend in the hospital last year?"

Table 3 contains descriptive statistics for the variables we use in the analysis. The first set of rows contains the outcome variables. Eligible individuals visit the doctor slightly more often, and report to be in slightly better health. They report to be more likely to stay in a hospital and to spend more nights in a hospital on average. Privately insured individuals are less likely to see

[13]Questions in this section are again taken from the English translation of the 2006 GSOEP personal questionnaire.

a doctor at all, and have on average less doctor visits. However, given that they see a doctor they see it more often. They are less likely to spend at least one night in a hospital. However, they spend more nights on average.

The second set of rows contains individual characteristics. Gross income is, by construction, on average higher for eligibles. In light of this of this it is not surprising that it is higher for privately insured (because only those with high enough incomes are eligible to buy private insurance). The remaining rows are informative about selection into private insurance. Given the characteristics of public and private insurance it is relatively more attractive to buy private insurance for individuals without children and individuals that are not married. This is because children and spouses whose income is relatively low are automatically covered by the insurance of the individual. This is reflected in the summary statistics. Privately insured individuals have fewer children and less of them are married. They are, however, about equally old.

In this paper we exploit variation in income around the compulsory insurance threshold to identify the effect of private insurance on doctor visits and health. For this to be valid we need that the variation in income around the compulsory insurance threshold is random. This would be violated if individuals could influence their true income (not only the reported one) in order to become eligible to buy private insurance. Even if there is classical measurement error a histogram of income minus the compulsory threshold should show a spike to the right of zero because there is only something to gain if income is *higher* than the compulsory threshold (Imbens & Lemieux, 2008; Lee & Lemieux, 2009; Van der Klaauw, 2009). Figure 4 shows several histograms, with varying bin width, and neither of them provides substantive grounds for this. A formal test that is suggested by McCrary (2008) confirms this ($p = 0.000$).

Covariates, when they have an impact on the respective outcome of interest, need to be included if their distribution conditional on z_i changes when we move from values of z_i slightly below zero to values slightly above zero (Imbens & Lemieux, 2008; Lee & Lemieux, 2009; Van der Klaauw, 2009). Otherwise, they can be left out. We investigate this for the indicator for being

	(1) Ineligible	(2) Eligible	(3) Public insurance	(4) Private insurance	(5) Total
At least 1 doctor visit	0.553	0.557	0.557	0.524	0.554
	-	-	-	-	-
Doctor visits given at least one visit	3.005	2.962	2.976	3.128	2.987
	(3.460)	(3.383)	(3.395)	(3.829)	(3.428)
Doctor visits	1.662	1.650	1.657	1.641	1.656
	(2.977)	(2.922)	(2.934)	(3.182)	(2.953)
Self-assessed health	2.439	2.387	2.426	2.322	2.418
	(0.803)	(0.792)	(0.798)	(0.802)	(0.799)
At least 1 night in hospital	0.0701	0.0707	0.0718	0.0510	0.0703
	-	-	-	-	-
Nights in hospital	0.668	0.750	0.693	0.818	0.702
	(3.837)	(4.775)	(4.113)	(5.690)	(4.253)
Gross income	34709.7	52014.0	41117.0	51291.4	41886.9
	(5125.1)	(7693.9)	(10340.2)	(9237.2)	(10607.3)
At least one child	0.512	0.602	0.563	0.392	0.550
	-	-	-	-	-
Number of children	0.896	1.116	1.019	0.610	0.988
	(1.025)	(1.106)	(1.071)	(0.903)	(1.065)
Years of education	11.46	13.13	12.02	13.77	12.15
	(2.395)	(3.019)	(2.676)	(3.617)	(2.796)
Married	0.703	0.785	0.751	0.563	0.737
	-	-	-	-	-
Age	40.89	42.37	41.55	41.04	41.51
	(7.633)	(6.769)	(7.352)	(6.853)	(7.316)
N	3797	2691	5997	491	6488

Means and standard errors (between parentheses). For binary variables only proportions are shown.

Table 3: Descriptive statistics

married, years of education and age. The second and third column of Table 3 already suggest that there is no need to include covariates because the respective means are very similar. This is supported by formal tests. For this we regress, locally using a rule-of-thumb bandwidth, the respective covariate on polynomials in z_i, one to the left and one to the right of the discontinuity. Then we test whether the respective predicted values at the discontinuity are equal. This is done for polynomials of order one (local linear), two (local quadratic), and three (local cubic). The results in Table 4 show that for none of the variable there is a discontinuity at the threshold.

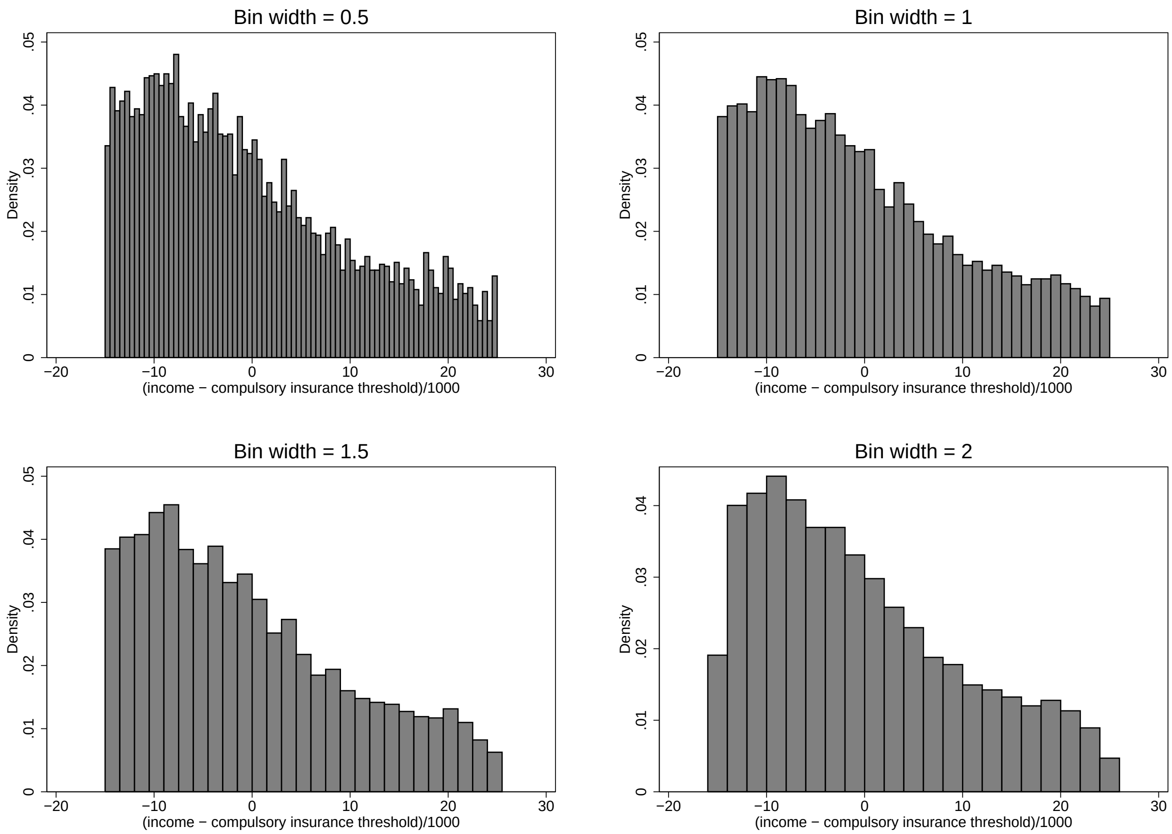

Figure 4: Histogram for the forcing variable

order of polynomial	1	2	3
Married	-0.008	-0.018	-0.032
	(0.064)	(0.072)	(0.099)
Years of education	-0.252	-0.230	-0.070
	(0.389)	(0.429)	(0.587)
Age	-0.682	-0.589	-1.140
	(1.037)	(1.214)	(1.491)

Standard errors are shown between parentheses.

Table 4: Local polynomial test for discontinuity at the threshold for different covariates

6 Results

We jointly estimate an equation for the probability to be privately insured conditional on reported income,

$$\mathbb{E}\left(p_i|z_i\right) = \Phi\left(\frac{z_i}{\sigma_u}\right) \cdot \left(\alpha_t + \beta z_i + \beta\sigma_u \frac{\phi\left(\frac{z_i}{\sigma_u}\right)}{\Phi\left(\frac{z_i}{\sigma_u}\right)}\right), \tag{4}$$

and an equation for medical care utilization conditional on reported income

$$\begin{aligned}\mathbb{E}\left(y_i|z_i\right) = &\left(1-\Phi\left(\frac{z_i}{\sigma_u}\right)\right)\left(\alpha_0 + \beta_0 z_i - \beta_0\sigma_u \frac{\phi\left(\frac{z_i}{\sigma_u}\right)}{1-\Phi\left(\frac{z_i}{\sigma_u}\right)}\right) \\ &+ \Phi\left(\frac{z_i}{\sigma_u}\right)\left(\alpha_0 + \Delta^{LATE}\cdot\alpha_t + \beta_1 z_i + \beta_1\sigma_u \frac{\phi\left(\frac{z_i}{\sigma_u}\right)}{\Phi\left(\frac{z_i}{\sigma_u}\right)}\right),\end{aligned} \tag{5}$$

where α_t is a year specific jump, at the compulsory insurance threshold, in the probability to be privately insured. This is reasonable since the compulsory insurance threshold changed over time (see Table 1). Here we impose that the local average treatment effect is the same in all years, i.e. we impose that Δ^{LATE}, our parameter of main interest, is independent of z_i^*. Then, it follows from equation (3) that we can replace α_1 by $\alpha_0 + \Delta^{LATE}\cdot\alpha_t$. Here, the size of both the numerator the and denominator in equation (1) is still allowed to vary across years, but we impose that the relative change in both is the same.

We first estimate equation (4) alone. Estimation results are provided in Table 5.[14] Coefficient estimates are marginal effects because this is a linear probability model, where the probability is linear in the unobserved z_i^*. However, it is nonlinear in z_i, see equation (4). The results indicate that for all years there is a discontinuous jump in the probability to buy private insurance at $z_i^* = 0$. It is significant for all years but 2000. Between 2001 and 2003 the jump is about 5

[14]Estimates were very similar when we estimated equation (4) and (5) together.

	(1)
(Gross income - threshold)/10000	0.087***
	(0.010)
Discontinuity 2000 (α_{2000})	0.031
	(0.020)
Discontinuity 2001 (α_{2001})	0.048**
	(0.021)
Discontinuity 2002 (α_{2002})	0.057***
	(0.018)
Discontinuity 2003 (α_{2003})	0.047***
	(0.018)
Discontinuity 2004 (α_{2004})	0.116***
	(0.019)
Discontinuity 2005 (α_{2005})	0.122***
	(0.020)
Discontinuity 2006 (α_{2006})	0.105***
	(0.020)
σ_u^2	0.620***
	(0.109)
R^2	0.166
N	6488

Standard errors are clustered at the individual level and shown between parentheses. $*, **, ***$ denote significance at the 10, 5, and 1 % level, respectively.

Table 5: Estimation results of the probability being privately insured

	(1) At least one doctor visits	(2) Doctor visits for subsample	(3) Doctor visits	(4) Self-assessed health
Δ^{LATE}	-0.054	-4.598**	-2.574*	-0.755*
	(0.214)	(2.033)	(1.393)	(0.432)
Baseline outcome	0.556***	3.088***	1.729***	2.444***
	(0.011)	(0.086)	(0.061)	(0.021)

Standard errors are clustered at the individual level and shown between parentheses. $*, **, ***$ denote significance at the 10, 5, and 1 % level, respectively.

Table 6: Main results

percentage points. Between 2004 and 2006 the jump more than doubles to approximately 12 percentage points. Supposedly, this is due to the increase in the compulsory insurance threshold for income earned in 2003, which affects the probability to be privately insured in 2004. Equality of all jumps is rejected ($p = 0.000$).

Table 6 presents the estimates of Δ^{LATE} for doctor visits in the past three months and self-assessed health. The respective baseline outcome is the average outcome for publicly insured individuals with z_i^* just below zero.

In specification (1) we use an indicator for at least one doctor visit as the dependent variable. This is a linear probability model since the expected outcome is a probability. 55 per cent of the publicly insured individuals see a doctor at least once within three months. We find no significant effect of private insurance on this. Manning, Morris, and Newhouse (1981) point out that the decision to visit a doctor at all, the so-called contact decision, is made by the individual, whereas the number of visits is mainly determined by the doctor who determines the intensity of the treatment.[15] In light of this it seems reasonable that we don't find an effect of private insurance on the probability to see the doctor at least once, as the decision to visit a doctor can be thought of as being motivated by some illness that is exogenous to the individual.

In specification (2) we also estimate the effect of private insurance on the number of doctor visits for those individuals who visit a doctor at least once. The baseline outcome is 3.088. The effect of private insurance on this is estimated to be negative and significant. The estimated magnitude of the effect, however, is too big. Nevertheless, the 95 per cent confidence interval ranges from -8.583 to -0.613, and the upper end seems reasonable in terms of the magnitude.

Specification (3) is for the number of doctor visits in the entire sample. This is a combination of the two effects we discussed above. The mean baseline outcome is estimated to be 1.729. Again, the effect of private insurance is negative and significant, now at the 10 per cent level, but

[15]Using GSOEP data Pohlmeier and Ulrich (1995) and Jürges (2009) use a count data hurdle model to relate the contact and frequency decision to covariates. One of the covariates in Jürges (2009) is whether the individual is privately insured. He does, however, not control for selection into private insurance.

again the magnitude of the point estimate of the effect seems to be too big.

Finally, we find a positive effects on health, recalling that a negative coefficient on insurance status means a positive effect. The estimate is significant at the 10 per cent level. This suggests that private patients receive better treatment despite going to the doctor less often overall. Importantly, our estimation procedure controls for selection into private insurance, so this could not be due to unobserved differences between publicly and privately insured individuals.

7 Sensitivity analysis

Our approach to correct for measurement error in the income variable is parametric because we have made both functional form and distributional assumptions for this. We run the risk of obtaining biased estimates if these are violated. Moreover, we have assumed that the measurement error is in the income measure, and not in insurance status. In this section we show that our main conclusions are robust to relaxing these assumptions.

For this we estimate the difference in the respective outcome between individuals with reported values of z_i slightly above zero and slightly below zero. We do so by regressing, locally using a rule-of-thumb bandwidth, the respective outcome on polynomials in z_i, one to the left and one to the right of the discontinuity. Then we calculate the difference between the predicted outcome, at the discontinuity, when we use the polynomial that is estimated to the right minus the predicted outcome using the polynomial that is estimated to the left of the discontinuity. This is done for polynomials of order one (local linear), two (local quadratic), and three (local cubic).

Battistin et al. (forthcoming) show that under the assumption that at least some individuals report their income accurately, which seems plausible, this is lower bound on the magnitude of the numerator in (1).[16] Moreover, and more importantly, they also show that the sign of the

[16] Before we have assumed that the measurement error is normally distributed with some positive variance. Strictly speaking, this implies that with probability zero income is reported accurately. However, a positive fraction will, under this assumption, record their income almost accurately in the sense of a deviation that is no bigger than a certain amount, say e.g. 300 Euros per year.

	Order 1		Order 2		Order 3	
	estimate	s.e.	estimate	s.e.	estimate	s.e.
At least 1 doctor visit	0.007	(0.067)	0.014	(0.081)	0.034	(0.101)
Doctor visits given at least 1 visit	-1.206	(0.758)	-1.645*	(0.962)	-2.115*	(1.162)
Doctor visits	-0.385	(0.460)	-0.864	(0.614)	-1.015	(0.716)
Self-assessed health	-0.032	(0.109)	-0.095	(0.141)	-0.066	(0.181)

Standard errors are clustered at the individual level and shown between parentheses. $*, **, ***$ denote significance at the 10, 5, and 1 % level, respectively.

Table 7: Local polynomial estimates of the discontinuity at the threshold

estimate of the numerator is equal to the sign of the effect.

Table 7 reports the results for the outcomes. Our main findings in this paper are that private insurance has a negative effect on doctor visits and a positive effect on health. The table shows that the sign is always estimated to be negative for doctor visits, also if we condition on at least one doctor visit. Similarly, the sign is the same as before for health (recall that a negative sign means a positive effect). The estimates of the effects, however, are not significant. This could be due to the measurement error in income, which drives them to zero (attenuation bias). Nevertheless, the table demonstrates that the point estimate of the sign of the effect is robust to relaxing the distributional and functional form assumptions.

The same holds true for the sign of the effect if in fact income is accurately measured but insurance status is measured with error. This follows because then the denominator in (1) can never be negative (given the institutions) and estimation of the denominator in (1) does not involve insurance status.

Finally, we re-do the analysis for the probability of at least on night spent in a hospital and the number of nights spent in a hospital. Table 8 shows the results. We find no effects of private insurance, confirming the results of Geil et al. (1997) and Riphahn et al. (2003).

Apart from these sensitivity analyses we have conducted further robustness checks. In the

	At least 1 night spent in a hospital	Nights spent in a hospital
Δ^{LATE}	0.050	-1.071
	(0.107)	(1.553)
Baseline outcome	0.069***	0.735***
	(0.005)	(0.081)

Standard errors are clustered at the individual level and shown between parentheses.
$*, **, ***$ denote significance at the 10, 5, and 1 % level, respectively.

Table 8: Results for nights spent in a hospital

main analysis we have assumed that the treatment effect is the same across years. Moreover, we have imposed linearity of $\mathrm{E}(y_i|z_i^*)$ and $\mathrm{E}(p_i|z_i^*)$ in z_i^* (not z_i) to the left and to the right of the discontinuity, respectively. These assumptions have been relaxed as part of a sensitivity analysis. In particular, we have estimated an unrestricted model and have used a piecewise linear specification. Findings of this suggest that our baseline specification is appropriate and because the results add little information they are not presented here.

8 Conclusions

In this paper we estimate the effect of private health insurance on the number of doctor visits and self-assessed health in Germany. Variation in income around the compulsory insurance threshold generates a natural experiment which allows us to control for selection into private insurance and estimate respective average treatment effects for individuals that would opt out for private insurance once becoming eligible. We show that it is important to account for measurement error in income and suggest a way to do so.

We find a significant negative effect of private insurance on doctor visits, at least for those individuals who see the doctor at least once. At the same time we find significant and sizable positive effects of private health insurance on satisfaction with health. This suggests that privately insured individuals receive better treatment when going to the doctor.

Appendix A : Derivations

In this appendix we derive an expression for $\mathbb{E}(p_i|z_i^*) = \Pr(p_i = 1|z_i^*)$. Recall that $z_i = z_i^* + u_i$, where u_i is normally distributed with mean 0 and variance σ_u^2, statistically independent of z_i^*, p_i and the potential outcomes. For $z_i^* < 0$ we have that $\mathbb{E}(p_i|z_i^*) = 0$ by definition. For $z_i^* \geq 0$ we specify $\mathbb{E}(p_i|z_i^*)$ to be a linear function in z_i^*, a linear probability model. That is,

$$\mathbb{E}(p_i|z_i^*) = \begin{cases} 0 & \text{if}\, z_i^* < 0 \\ \alpha + \beta z_i^* & \text{if}\, z_i^* \geq 0. \end{cases}$$

By the law of total probability,

$$\mathbb{E}(p_i|z_i) = \Pr(z_i^* < 0|z_i) \cdot 0 + \Pr(z_i^* \geq 0|z_i) \cdot \mathbb{E}(p_i|z_i, z_i^* \geq 0).$$

The assumptions about the measurement error imply that this is equivalent to

$$\mathbb{E}(p_i|z_i) = \Pr(u_i \leq z_i) \cdot (\alpha + \beta \mathbb{E}(z_i - u_i|z_i, u_i \leq z_i)). \tag{6}$$

Recall that if v is standard normally distributed then $\mathbb{E}(v|v < c) = -\phi(c)/\Phi(c)$, which is known as the inverse Mills ratio, where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal cumulative distribution function and the probability density function, respectively. Using this equation (6) can be rewritten as

$$\mathbb{E}(p_i|z_i) = \Phi\left(\frac{z_i}{\sigma_u}\right) \cdot \left(\alpha + \beta z_i + \beta \sigma_u \frac{\phi\left(\frac{z_i}{\sigma_u}\right)}{\Phi\left(\frac{z_i}{\sigma_u}\right)}\right).$$

References

Akerlof, G. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, *84*, 488-500.

Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, *53*, 941-973.

Battistin, E., Brugiavini, A., Rettore, E., & Weber, G. (forthcoming). The retirement consumption puzzle: Evidence from a regression discontinuity approach. *American Economic Review*.

Battistin, E., & Rettore, E. (2008). Ineligibles and eligible non-participants as a double comparison group in regression discontinuity designs. *Journal of Econometrics*, *142*, 715-730.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In J. Heckman & E. Leamer (Eds.), *Handbook of Econometrics, volume 5* (p. 3705 - 3843). Elsevier Science.

Buchmüller, T., Grumbach, K., Kronick, R., & Kahn, J. (2005). The effect of health insurance on medical care utilization and implications for insurance expansion: A review of the literature. *Medical Care Research and Review*, *62*, 3 - 30.

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

Card, D., Dobkin, C., & Maestas, N. (2009). Does Medicare save lives? *Quarterly Journal of Economics*, *124*, 597 - 636.

Chiappori, P., Durand, F., & Geoffard, P. (1998). Moral hazard and the demand for physician services: First lessons from a French natural experiment. *European Economic Review*, *42*, 499 - 51.

Colombo, F., & Tapay, N. (2004). *Private health insurance in OECD countries: The benefits and costs for individuals and health systems.* OECD Health Working Paper No 15.

Cutler, D., & Zeckhauser, R. (2000). The anatomy of health insurance. In J. Newhouse & A. Culyer (Eds.), *Handbook of Health Economics, Volume 1* (p. 563 - 643). Elsevier Science.

Deveugele, M., Derese, A., van den Brink-Muinen, A., Bensing, J., & De Maeseneer, J. (2002). Consultation length in general practice: Cross sectional study in six european countries. *British Medical Journal*, *325*, 472-477.

Fang, H., Keane, M., & Silverman, D. (2008). Sources of advantageous selection: Evidence from the medigap insurance market. *Journal of Political Economy*, *116*, 303 - 350.

Finkelstein, A., & McGarry, K. (2006). Multiple demensions of private information: Evidence from the long-term care insurance market. *American Economic Review*, *96*, 938 - 958.

Geil, P., Million, A., Rotte, R., & Zimmerman, K. (1997). Economic incentives and hospitalization in Germany. *Journal of Applied Econometrics*, *12*, 295-311.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression discontinuity design. *Econometrica*, *69*, 201 - 209.

Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*, 467-475.

Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to pratice. *Journal of Econometrics*, *142*, 615 - 635.
Jürges, H. (2009). Health insurance status and physician behavior in Germany. In (Vol. 129, p. 297 - 307).
Kreider, B. (forthcoming). Regression coefficient identification decay in the presence of infrequent classification errors. *Review of Economics and Statistics*.
Kreider, B., & Hill, S. (forthcoming). Partially identifying treatment effects with an application to covering the uninsured. *Journal of Human Resources*.
Lee, D., & Lemieux, T. (2009). *Regression discontinuity designs in economics.* Working paper No 14723, NBER.
Levy, H., & Meltzer, D. (2004). What do we really know about whether health insurance affects health. In C. McLaughlin (Ed.), *Health Policy and the Uninsured.*
Lungen, M., Stollenwerk, B., Messner, P., Lauterbach, K. W., & Gerber, A. (2008). Waiting times for elective treatments according to insurance status: A randomized empirical study in germany. *International Journal for Equity in Health*, *7*(1), 1-7.
Manning, W., Morris, C., & Newhouse, J. (1981). A two-part model of the demand for medical care: preliminary results from the health insurance study. In J. van der Gaag & M. Perlman (Eds.), *Economics and health economics.* Amsterdam: North Holland.
Manning, W., Newhouse, J., Duan, N., Keeler, E., Leibowitz, A., & Marquis, M. (1987). Health insurance and the demand for medical care: evidence from a randomized experiment. *American Economic Review*, *77*, 251 - 277.
McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*, 698 - 714.
Newhouse, J. (1974). A design for a health insurance experimemt. *Inquiry*, *11*, 5 - 27.
Newhouse, J. (1993). *Free for all? lessons from the rand health insurance experiment*. Harvard University Press.
Pohlmeier, W., & Ulrich, V. (1995). An econometric model of the two-part decisionmaking process in the demand for health care. *Journal of Human Resources*, *30*(2), 339-361.
Riphahn, R., Wambach, A., & Million, A. (2003). Incentive effects in the demand for health care: A bivariate panel count data estimation. *Journal of Applied Econometrics*, *18*, 387 - 405.
Rothschild, M., & Stiglitz, J. (1976). Equilibrium in competitive insurance markets: an essay in the economics of incomplete information. *Quarterly Journal of Economics*, *90*, 624-649.
Thistlethwaite, D., & Campbell, D. (1960). Regression discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, *51*, 309 - 317.
Van der Klaauw, W. (2009). Regression discontinuity anlaysis: A survey of recent developments in economics. *Labour*, *22*, 219 - 245.