

Schmitz, Hendrik

Conference Paper

Practice budgets and the patient mix of physicians - Evaluating the effects of remuneration system reforms on physician behaviour in Germany

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session:
Health Policy: Patients, Doctors, Hospitals, No. D18-V2

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Schmitz, Hendrik (2010) : Practice budgets and the patient mix of physicians
- Evaluating the effects of remuneration system reforms on physician behaviour in Germany,
Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Health
Policy: Patients, Doctors, Hospitals, No. D18-V2, Verein für Socialpolitik, Frankfurt a. M.

This Version is available at:

<https://hdl.handle.net/10419/37176>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen
Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle
Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich
machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen
(insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten,
gelten abweichend von diesen Nutzungsbedingungen die in der dort
genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

*Documents in EconStor may be saved and copied for your personal
and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to
exhibit the documents publicly, to make them publicly available on the
internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content
Licence (especially Creative Commons Licences), you may exercise
further usage rights as specified in the indicated licence.*

Practice budgets and the patient mix of physicians

- Evaluating the effects of remuneration system reforms on physician behaviour in Germany

Hendrik Schmitz*

Ruhr Graduate School in Economics (RGS Econ)
and University of Duisburg-Essen

November 12, 2009

Abstract

We analyse the effect of a change in the remuneration system for physicians on the lengths of treatments of patients as measured by the number of doctor visits using data from the German Socio-Economic Panel over the period 1988-2003. In particular, we analyse the introduction of a fixed budget for the ambulatory sector in 1993 and the introduction of individual practice budgets in 1997 for the publicly insured. With a random-effects-type two-part model, we find evidence that the reforms did not change the patients' behaviour (and access to health care) but that the introduction of individual practice budgets in 1997 reduced the treatment durations of the publicly insured. At the same time, treatment durations increased for the privately insured. This can be seen as evidence that physicians respond to the change in incentives induced by the reform.

JEL Classification: I11, I18

Keywords: supplier-induced demand, latent class hurdle model, ambulatory care, practice budgets

*I am grateful to Leilanie Basilio, Daniel Baumgarten, Stefanie Neimann, Alfredo Paloyo, Reinhold Schnabel, Nicolas Ziebarth, and referees for valuable comments. This paper benefited from the discussions at the RGS Workshop in Essen and the Econometric Society European Meeting in Barcelona 2009. Financial support by the Leibniz Association is gratefully acknowledged. All correspondence to Hendrik Schmitz, Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Hohenzollernstr. 1-3, 45128 Essen, Germany, Fax: +49-201-8149-500; E-mail: hendrik.schmitz@rwi-essen.de

1 Introduction

Steadily increasing health care costs have been an issue in most industrialised countries for the last few decades. In Germany, however, health care expenditures as a fraction of the gross domestic product have been rather stable since the mid-1990s even though demographic change and technological progress have increased the pressure on the health care system. For instance, the share was 10.4% in 1996 and it stood at 10.6% by 2006. This fraction has risen by 12 to 20% in countries like Switzerland, France, the US, and the UK in the same period (OECD, 2008). A major reason for the stability in Germany could be the introduction of fixed budgets for various health care sectors (stationary, ambulatory, and pharmaceuticals) that increase only by a limited amount each year. These fixed budgets are in general a means to introduce rationing in the health care market.

The budgets in the ambulatory sector directly affect the remuneration of physicians and, therefore, possibly also their behaviour. This study analyses the effect of the introduction of fixed budgets on physician's behaviour as measured by the length of treatment of patients. It uses a particularity of the German health insurance system, namely the existence of two different and independent insurance systems (private and statutory public insurance) that imply different remuneration systems for physicians. Until 1993, the remuneration for treating the statutorily insured was based on a fee-for-service (FFS) system. In 1993, the remuneration system was reformed with the introduction of a fixed budget and a point system for the statutorily insured. From then on, physicians got points for each treatment. The monetary value of each point was then calculated at the end of each quarter by dividing the total budget by the sum of all points collected by all physicians. In 1997, a further reform was introduced which capped the total points reported by doctors by a so-called "individual practice budget". Since no reform took place in the private sector in the period of analysis, this allows us to analyse the response of physicians to the change in financial incentives by using the reform as a source of exogenous variation.

This study contributes to a growing literature that analyses the effect of physician remuneration on the quantity of health care utilisation, typically measured by the number of physician visits. As found by Devlin and Sarma (2008), physicians conduct more patient visits under FFS than under any other remuneration system in Canada. Croxson et al. (2001) and Dusheiko et al. (2006) find effects of budgets on physician behaviour for the UK. Hennig-Schmidt et al. (2008) show in an experimental setting that physicians respond to incentives imposed by the reimbursement system and that they tend to overtreat patients under FFS and to undertreat them

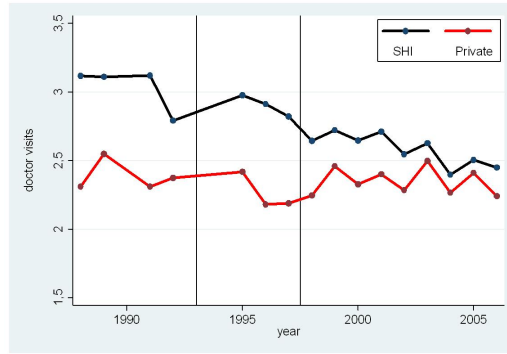
in a per-capita payment system. On the other hand, Madden et al. (2005) and Grytten and Sorensen (2001) do not find significant effects of the remuneration system on physician behaviour for Ireland and Norway. Implicitly, studies on the effects of the remuneration system on physicians' behaviour can be seen as analyses of supplier-induced demand. This is true at least for remuneration systems like the FFS-system, which provide incentives for physicians to conduct excess treatments.

To our knowledge, this is the first paper that evaluates the impact of the introduction of the fixed budget and the individual practice budget in Germany (see Wörz and Busse, 2005, who also note the absence of any scientific evaluation of these reforms). Up to now, German literature that analysed supplier-induced demand has concentrated on the effect of physician density on the number of doctor visits and findings show only weak and mixed evidence for Germany (Krämer, 1981; Breyer, 1984; Pohlmeier and Ulrich, 1995; Kopetsch, 2007). Only Jürges (2009) explicitly accounts for the differences in the remuneration system and finds that, in the year 2002, those who were privately insured had, on average, more doctor visits given that they had contacted a doctor. Moreover, while physician density increased the frequency of doctor visits for all patients, the effect was strongest for the privately insured.

Figure 1 illustrates the evolution of the average number of doctor visits in the last three months prior to the interview using data from the German Socio-Economic Panel (SOEP) between 1988 and 2006. While the average number for the privately insured (around 10% of the population in 2006) stayed fairly constant around 2.3, the number for the publicly insured (the remaining 90%) steadily decreased from about 3.1 in 1988 to 2.5 in 2006. However, as Figure 2a shows, the probability of at least one visit slightly increased in the same period for both groups while the number of doctor visits for those individuals who had seen a doctor at least once in the previous three months (Figure 2b) decreased sharply for the publicly insured (from 4.7 to 3.5) and only slightly for the privately insured (from 3.8 to 3.5). Hence, the decline in the average number of doctor visits for the statutorily insured is almost exclusively a result of the decline in the number of visits for those individuals who had at least one doctor visit.

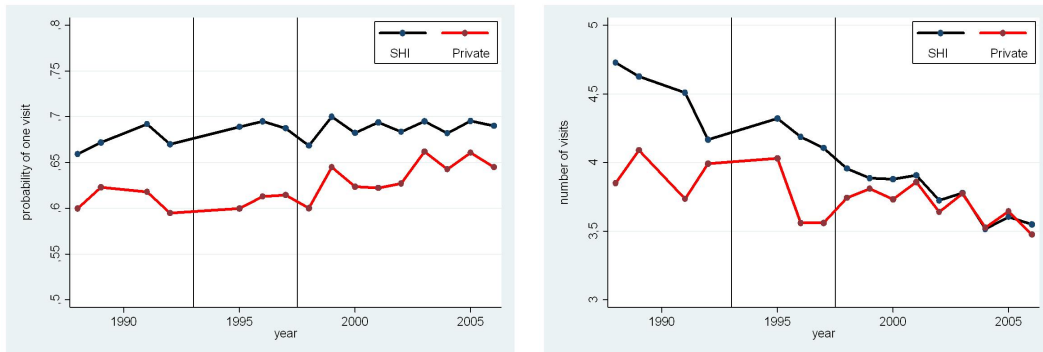
The slight increase in the probability of at least one doctor visit (Figure 2a) might reflect the growing importance of preventive doctor visits while the observed decline in the total number of doctor visits (Figure 1 and Figure 2b) may have several reasons. First, because this is a long panel, composition effects might play a role. That is, the panel might have changed in observables like age, education, or the health status of the respondents. All of these variables are important for the demand for doctor visits. Second, panel attrition could matter as well. It can

Figure 1: Average number of doctor visits in previous three months, overall



Source: SOEP, years 1988-2006. Vertical lines represent the years 1993 and 1997.

Figure 2: Average ... in previous three months



(a) ... probability of at least one doctor visit (b) ... number of doctor visits, conditional on one visit

Source: SOEP, years 1988-2006. Vertical lines represent the years 1993 and 1997.

be expected that unhealthy individuals (with a high demand for doctor visits) have a higher probability to drop out of the data set due to severe illness, death, or other reasons. Frijters et al. (2005) show that panel attrition of individuals with lower health satisfaction is an issue in the SOEP; Contoyannis et al. (2004) find similar problems in the British Household Panel Study (BHPS). These two points - compositional changes in observables and unobservables - could lead to a decline in the average number of doctor visits in the SOEP. However, because the SOEP is a representative panel that saw several refreshments in the observation period, both points should not be the only reasons for this picture.

Third, macroeconomic factors like the unemployment rate contribute to the demand for doctor visits. As is well established (e.g., Askildsen et al., 2005), work absenteeism is less frequent in recessions due to a greater fear of losing the job. Workers who are absent for more than three days, however, need a certificate from a doctor and, thus, a doctor visit. Hence, recessions could have decreased the number of doctor visits. Fourth, institutional changes concerning the supply side of health

services (here, the physicians as providers of outpatient care) might have had an impact.

This study takes into account all four points but puts emphasis on the last one. It turns out that, after controlling for compositional effects, panel attrition, and economic conditions, the decline in the number of visits of the publicly insured is much less pronounced and can be attributed to the introduction of individual practice budgets in 1997. Moreover, not only did the number of doctor visits of the publicly insured decrease after the reform, it also increased for the privately insured. This gives rise to the interpretation that physicians responded to the reforms by changing their patient mix, i.e., by substituting out the publicly insured for the privately insured. The results are robust to different specifications. Moreover, several tests support the identifying assumptions, one of these being the test for the single spell assumption as derived by Santos Silva and Windmeijer (2001).

The remainder of this paper is structured as follows. Section 2 gives an overview of the institutional background and of the reforms that took place in the German health care system. Section 3 explains the data, Section 4 the empirical strategy. Section 5 presents the estimation results. Section 6 shows the robustness of the results and supports the identifying assumptions while Section 7 concludes.

2 Payment System and Major Reforms

The German health insurance system consists of two parts. About 90% of the population are insured by statutory health insurance (SHI; also called public insurance hereafter). It is compulsory for all individuals with earnings below a certain income threshold (3,975 Euro per month in 2007) and who are not civil servants or self-employed. It is financed by payroll taxes and non-working family members are covered without an extra premium. The benefit package is heavily regulated and does not vary much between insurance companies. Individuals who earn more than the income threshold, the self-employed, and civil servants are allowed to opt out of the statutory insurance system and instead buy private insurance. This group accounts for the remaining 10% of the German population. The private insurance premium does not depend on income but instead is a risk-equivalent contribution depending on age, gender, and health status when the contract is signed. Privately insured individuals have to pay higher premia in order to cover non-working family members. Thus, having many dependents is a reason for staying voluntarily in the public system for about 50% of all the individuals who are eligible to opt out.

Physicians are remunerated according to an FFS-system. Before 1993, the price for a treatment was fixed *ex ante* and depended on the complexity of the treatment.

Treatments of the statutorily insured were (and still are) charged according to the EBM (“Einheitlicher Bewertungsmaßstab”), whereas treatments of privately insured were charged according to a different legal setting, namely the GOÄ (“Gebührenordnung für Ärzte”). The statutory health insurance is a full cover insurance (with some exceptions). The insurance company directly pays for the treatments and hence the patient does not see the costs she actually incurs. In such a system, neither the patient nor the doctor has an incentive to contain costs (Jürges, 2009). On the contrary, due to the well known information asymmetry between patient and physician, the physician can possibly induce demand to increase income. This incentive system might have contributed to an average number of doctor visits in Germany that is higher than in most other countries.

In reaction to the steadily increasing health care expenditures, the German government introduced a fixed budget for ambulatory fees for the statutorily insured in 1993. Under this system, doctors receive points for each treatment according to the severity of the case. At the end of each quarter, the monetary value for each point is calculated as the value of the total budget divided by the sum of all points collected by all doctors. The budgets and the sum of all points are determined regionally. Hence, the monetary value of one point varies by region and time.

The fixed budget was introduced in order to keep the overall costs stable for the social health insurance system. However, it cannot contribute to a reduction of medical services by physicians, as Benstetter and Wambach (2006) show theoretically. For instance, a single physician can still increase her income by increasing the duration of treatments. Given the fixed budget, however, and no coordination between physicians, this can only be at the cost of the point value. A decreasing point value again leads to increased activity of the physician, resulting in a further decrease in the point value. This is called the “treadmill effect”. Indeed, the point value declined after 1993 (Benstetter and Wambach, 2006). This, however, was not due to an increasing number of physician visits but due to the fact that physicians charged much more services during a treatment, especially doctor’s advice. The number of consultations even decreased by 6.5% but the number of charged points increased by more than 30% (Witteck, 1996).¹

Although the increasing dispensation of doctor’s advice was partly intended by the policy makers to strengthen the “speaking medicine”, some groups of physicians were negatively affected by the reform while others benefited.² In order to stop the

¹All numbers are only for the statutorily insured. Therefore, the decrease in doctor visits need not result from the reform but – due to the absence of a control group – can be due to a temporary shock. In fact, below, we evaluate the reform effects on the number of consultations including a group that was not directly affected, the privately insured.

²The net effect for all physicians was zero in monetary terms since the total budget was fixed

Table 1: Payroll tax base and CPI, normalised

Year	2000	2001	2002	2003	2004	2005	2006
Payroll tax base	100.0	102.1	102.8	102.8	103.4	103.2	104.7
CPI	100.0	102.0	103.4	104.5	106.2	108.3	110.1

Source: RWI Essen, base year is 2000.

declining point value, individual practice budgets were introduced in addition to the overall fixed budget in July 1997. With this reform, each physician receives a maximum number of points she could reimburse for each quarter. Points exceeding the practice budget were reimbursed by a much lower point value. The reform was successful in stabilising the point value (see Benstetter and Wambach, 2006). Since 1999, the budget for ambulatory care may not rise faster than the payroll tax base in Germany. However, since 2000, the increase in the payroll tax base was lower than the inflation rate (see Table 1). Particularly in 2003, the budget did not increase at all. Altogether, this led to a reduction of the budget in real terms after 2000.

Given the incentives induced by the reform of 1993, a decrease in the number of doctor visits cannot be expected after the introduction of the fixed budget because, as regards incentives for doctors, there is no strong difference between a system where physicians get money values or point values for a treatment. In contrast, the reform in July 1997 should have an effect on the physician's behaviour. If the physician incurs a cost for each treatment, she should have an incentive not to exceed the individual amount of points she can get reimbursed. If it was before the case that physician-induced demand (i.e., longer treatments than necessary) increased the income of (some) physicians, the incentive to carry out unnecessary treatments was reduced after the reform. Another possibility is to postpone recalls of publicly insured patients into the next quarter.

There might also be an incentive to substitute out the publicly for the privately insured. Reducing the number of recalls of the publicly insured leads to higher available capacities for physicians to spend on the privately insured. Since treatment of the privately insured does not affect the practice budget, physicians might have an incentive to fill the gap by boosting the recalls of the privately insured. The tightening of the budget in real terms after 2000 should have further strengthened these incentives.

The private insurance system has not seen any reform during the whole period. That is, while the reimbursement system for statutorily insured patients has substantially changed, it remained unchanged for privately insured patients. A first

 at the expenditure level prior to the reform.

indicator of a positive effect of rationing in the SHI is the development of health insurance contributions in the public and the private system. Using data from the SOEP for the period 1984-2006, Grabka (2006) shows that, while the contributions nominally increased by about 130% in the SHI, they tripled in the private insurance system.

A major problem in identifying the reform effect is that, in this long time period, two other reforms affected the demand (i.e., patient) side of the market. This is especially important since the data we use here measure the number of doctor visits at the patient level. First, at the same time that the introduction of practice budgets became effective, there was a reform that increased the co-payments for prescription drugs by 6 DM (about 3 Euro).³ Winkelmann (2004a, 2004b, 2006) finds remarkable negative effects of the reform on the demand for doctor visits using the same data as this paper. However, as we will argue later in Section 5, it is more likely that these effects are due to the introduction of practice budgets and not due to the increased co-payments for prescription drugs. Second, a co-payment of 10 Euro was introduced in 2004 for the first doctor visit in each quarter. While Augurzky et al. (2006) and Schreyögg and Grabka (2008) do find significant demand-side effects in the SOEP, Farbmacher (2009) finds a reduction in the probability of visiting a physician which is small but significant.

3 Data

Our empirical analysis is based on the German Socio-Economic Panel (SOEP), a survey which started in 1984 with more than 12,000 individuals in West Germany and was extended to include East Germany in June 1990. There were several refreshments resulting in a sample size of more than 20,000 adult individuals living in about 13,000 households that participated in the SOEP survey in 2006 (see, e.g., Wagner et al., 2007).⁴ The SOEP includes questions about the total number of doctor visits within the last three months prior to the interview in all years except for 1990 and 1993. In the years 1984 to 1987 and 1994, the SOEP does not ask for the total number but differentiates between general practitioners (GPs) and various kinds of specialists. Because these questions differ from the one in all other years,

³The reform was partly abolished in 1999.

⁴The data used in this paper were extracted using the Add-On package PanelWhiz v2.0 (Nov 2007) for Stata. PanelWhiz was written by Dr. John P. Haisken-DeNew (john@panelwhiz.eu). The PanelWhiz generated DO file to retrieve the SOEP data used here and any Panelwhiz Plugins are available upon request. Any data or computational errors in this paper are my own. Haisken-DeNew and Hahn (2006) describe PanelWhiz in detail.

we drop the years 1984-1987 and 1994 from the analysis.⁵

In order to not confound our results with the possible effects of the co-payment reform in 2004, we also disregard information from the years after 2003. Finally, we drop all individuals of the year 1997 that were interviewed after June. Because the question refers to the number of doctor visits in the previous three months, one cannot see if the stated number of visits falls into the period before or after the reform for these observations. Since most of the interviews take place in the first months of each year we do not lose many observations.

The main drawback of the SOEP for this kind of analysis is that it only provides the total number of doctor visits per respondent in the last three months. There are no information on the number and duration of illness episodes that are captured by this value. Hence, the first observed count within the quarter might be the continuation of a previous illness episode instead of being the initiation of a new spell. Furthermore, several visits in a quarter might either result from one longer illness episode or from multiple short ones. Considering the identification strategy outlined in the next section, the latter issue might be particularly problematic. Pohlmeier and Ulrich (1995) argue that the period of three months in the SOEP is a good compromise in reducing the former problem (that gets smaller with longer periods) without letting the latter problem get too large (which is reduced in shorter periods). In the following, we assume that the observed number of doctor visits results from only one sickness spell that starts at the beginning of the quarter to which the question refers. Santos Silva and Windmeijer (2001) call this the “single spell assumption” and argue that the validity of this assumption critically depends on the length of the observation period. In fact, this assumption can be tested. In Section 6.2 we outline the test for the single spell assumption that was derived by Santos Silva and Windmeijer (2001) and report the results.

The SOEP has several important virtues and it is probably the only available data set to answer the questions above. First, it covers a long period that starts well before the first reform became effective and continues to go on. It includes many variables that affect the individual demand for health care like the health status and many other socio-economic variables. Due to its panel nature, unobserved heterogeneity of the individuals can also be taken into account. Finally, and most importantly, it includes a group (the privately insured) that was not directly affected by either of the reforms. Although the groups of privately and statutorily insured are somewhat different, assuming that the group differences would have stayed stable

⁵Including these years does not affect the results at all. Because only the total number of visits is asked in all remaining years, we cannot distinguish effects between general practitioners and specialists because, apart from the mentioned years, the type of physician is not specified in the data set. Since the reforms affected all types of physicians we consider this a minor problem.

over time without the reforms makes it possible to identify the reform effect on physician behaviour.

Several variables are included to control for the differences between the privately and the publicly insured. The main variables are those which control for the observable health status. These are satisfaction with own health (on an 11-point scale), the number of hospital stays in the previous year, age, and if the respondent is handicapped. *Sport* controls for different health behaviour. The average unemployment rate and a dummy for job absenteeism for more than three days in the last year captures macroeconomic aspects that affect the demand for doctor visits. A dummy for West Germany captures different regional behaviour as well as different infrastructure. Furthermore, two variables are included to capture the effects of population density. Finally, a few other socio-economic control variables are included. Since males and females exhibit considerable differences in their doctor visiting behaviour, we carry out separate regressions for both groups. Table A2 in the appendix explains the variables and reports sample means.

4 Empirical Strategy

4.1 Count Data Hurdle Model

The dependent variable (number of doctor visits) takes on only non-negative integer values. Therefore, it seems reasonable to use a count data model such as the negative binomial model that has the following probability density function (see, e.g., Cameron and Trivedi, 2005):

$$f(y_{it}|\mu, \alpha) = \frac{\Gamma(\alpha^{-1} + y_{it})}{\Gamma(\alpha^{-1})\Gamma(y_{it} + 1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu} \right)^{\alpha^{-1}} \left(\frac{\mu}{\mu + \alpha^{-1}} \right)^{y_{it}}$$

where $\mu = \exp(x'_{it}\beta)$ and α is the over-dispersion parameter.

It is often argued that the observed number of doctor visits is a result of two different (and maybe independent) decision-making processes. In case of an illness, the patient decides whether or not to see a physician (1st stage). Once a doctor is seen, however, the doctor – maybe together with the patient – determines the length of the treatment (2nd stage). Hence, a hurdle model (also called a two-part model) seems to be the most appropriate formulation in order to explain the observed number of doctor visits (Mullahy, 1986; Pohlmeier and Ulrich, 1995). The underlying economic structure is that of a principal-agent model. That is, the first stage should mainly capture demand-side effects, while the second stage should also capture supply-side effects.

The availability of panel data allows to account for unobserved heterogeneity. We follow Bago d’Uva (2006) and add a time-invariant random effect that affects both stages. This random effect is supposed to follow a discrete distribution that takes on a small number of mass points. There is an economic and a statistical motivation for this resulting finite mixture model. The economic motivation classifies individuals into a small number of latent classes, e.g., two, the “high users” and the “low users”, with different effects of covariates on the outcome variable. There is a debate in the literature on whether the standard hurdle model (with the differentiation between “users” and “non-users”) or the finite mixture negative binomial model (with the less restrictive differentiation between “high users” and “low users”) is better able to explain data on doctor visits (see, e.g., Deb and Trivedi (1997), Deb and Trivedi (2002), Jimenez-Martin et al. (2002)). The advantage of the model derived by Bago d’Uva (2006), which is also used by Bago d’Uva and Jones (2009), is to combine both previous models and to allow for latent classes but, at the same time, to maintain the principal-agent structure of the model.

The latent class hurdle model that, in its most general form, allows for slope heterogeneity, however, has very many parameters to estimate and is very data-demanding. Here, we keep the model parsimonious and restrict it to intercept heterogeneity and to different over-dispersion parameters (i.e., different values of α) for the different latent classes. Thus, the motivation for the resulting finite mixture model comes more from a statistical side, that is, the possibility to introduce a time-invariant individual effect θ_i^m without imposing too many distributional assumptions on the effect, except for some general random-effects assumptions

$$E[\theta_i^m | x_{it}] = 0; E(\theta_i) = \sum_{m=1}^M P(\theta_i^m) \theta_i^m = 0; \sum_{m=1}^M P(\theta_i^m) = 1, \forall m(m = 1, \dots, M)$$

where M is the total number of mass points and $P(\theta_i^m)$ is the probability of mass point θ_i^m . The density of the observed data is given by

$$g_m(y|x, \theta_i^m) = \begin{cases} f_{1m}(0|x, \theta_i^m) & \text{if } y = 0 \\ (1 - f_{1m}(0|x, \theta_i^m)) f_{2m}(y|x, \theta_i^m, y > 0) & \text{if } y > 0 \end{cases}$$

where

$$f_{1m}(0|x_{it}, \theta_i^m) = P(y_{it} = 0|x_{it}, \theta_i^m, \beta_1) = (\mu_{1m} + 1)^{-1}$$

and

$$f_{2m}(y_{it}|x_{it}, \theta_i^m, \beta_2; y_{it} > 0) = \frac{\Gamma(\alpha_m^{-1} + y_{it})}{\Gamma(\alpha_m^{-1})(\Gamma(y_{it} + 1))((1 + \alpha_m \mu_{2m})^{\alpha_m^{-1}} - 1)} \left(\frac{\mu_{2m}}{\mu_{2m} + \alpha_m^{-1}} \right)^{y_{it}}$$

and $\mu_{1m} = \exp(x'_{it}\beta_1 + \theta_i^m)$, $\mu_{2m} = \exp(x'_{it}\beta_2 + \theta_i^m)$.

The parameter vectors β_1 , β_2 , the heterogeneity terms α_m , and the locations and probabilities of the mass points are estimated jointly by maximizing the following likelihood function, where T_i denotes the number of years individual i is observed in the data set.⁶

$$L = \prod_{i=1}^N \sum_{m=1}^M P(\theta_i^m) \prod_{t=1}^{T_i} g_m(y_{it}|x_{it}, \theta_i^m).$$

4.2 Estimation Strategy

In this study, we estimate the effects of two reforms on the patient mix of physicians. The analysis is complicated by the fact that the data are on the patient and not on the physician level. Given the principal-agent structure of the estimation model outlined in the previous section, we measure the physicians' behaviour in the second stage of the hurdle model using data of patients. The fact that the reform directly affected only the treatment of the publicly and not the privately insured gives reason to evaluate the effect of introducing fixed budgets for doctors on the number of doctor visits of the publicly insured using the group of the privately insured as a control group. However, this would mean assuming the absence of any general equilibrium effects, i.e., of any effects the reform might have had on the private sector. The theoretical discussion in Section 2 and the raw numbers in Figures 1 and 2b support the notion that the privately insured are also (indirectly) affected by the reform, because it might be that doctors substitute out the publicly for the privately insured. Therefore, the privately insured are not well-suited as a control group and the estimated treatment effect of a difference-in-differences analysis is likely be biased.

Nevertheless, in the estimation, we include a dummy for being publicly insured, time dummies, and interactions between time dummies and the indicator for public insurance (see Table 2 for the structure of the time dummies). We do this to compare the expected number of doctor visits of the publicly insured with the one of the privately insured before and after the reform, keeping in mind that the difference

⁶We use the *ml* command in Stata and the Broyden-Fletcher-Goldfarb-Shanno quasi-Newton algorithm. We draw from the code provided in Jones et al. (2007). Different starting values are used to rule out local maxima of the likelihood function.

does not necessarily measure the effect of introducing budgets on the number of doctor visits of the publicly insured.

Table 2: Covariates

Year	Regulations	T_{93-97}	T_{98-99}	T_{00-03}
1988 - 1992	Base system	0	0	0
1993 - 1997	Fixed budget	1	0	0
1998 - 1999	Individual budget	0	1	0
2000 - 2003	Stricter individual budget	0	0	1

By doing this, we can analyse whether or not physicians changed their patient mix after the reform. Ideally, one would want to compare the patient mix of doctors who were affected by the reform with the one of doctors who were not affected. Since virtually all physicians treat both privately and publicly insured, there is no control group of physicians available and therefore no true difference-in-differences analysis possible.⁷ Our analysis is therefore a before-after evaluation. In order to identify a causal reform effect, we do not only have to assume absence of unobserved effects that differ for treatment and control groups (like in a difference-in-differences analysis) but absence of any unobserved effects that change the patient mix of doctors over time. This, however, does not seem to be a very restrictive assumption in our case. Assuming that there are no exogenous shocks that affect the *absolute number* of doctor visits might be too strong (think of flu epidemics, for example), making a reform evaluation without a control group unfeasible when the patient behaviour is the object of interest. However, we see no reason why unobserved effects should alter the *relative number* of visits, i.e., why doctors should change their patient mix for reasons other than the reform (i.e., privately and publicly insured are not affected differently by flu epidemics).

For the analysis to be valid, we furthermore have to assume common trends for publicly and privately insured. That is, we assume that the existing differences in the expected number of doctor visits between both groups would have stayed stable without the reform. Obviously, this counterfactual situation is not testable but we come back to this point in Section 6.1 (Robustness Checks). We also assume that individuals do not self-select into public or private insurance due to the reform. One might argue that publicly insured individuals who need many doctor visits realise that they would get shorter treatments due to the reform and therefore change

⁷There is a very small group of physicians that specialise on treating privately insured patients only (less than 1% of all physicians). However, this group is not only small but usually specialises in fields like psychotherapy and is therefore not well-suited as a control group.

into the private system. Because these would be the sicker individuals, the average number of doctor visits in the SHI would decrease and the number in the private system increase. This is very unlikely to be the case in Germany. First, only a small number of individuals (about 20%) can actually choose between statutory and private insurance. The remaining 80% cannot change even if they want to. Second, sick individuals in particular would decide to stay in the statutory system because private insurers are allowed to use the health status in calculating insurance premia. Thus, the contribution is higher in the private system than in the public system for sick individuals or, even worse, some services might be excluded from the private benefit package at the outset. However, as a robustness check, we carry out a regression including only those individuals who never switched health insurances after 1996.

4.3 Control for Panel Attrition

Although there were several refreshments in the SOEP, panel attrition might be a problem in such a long period of analysis (Frijters et al., 2005). It might be the case that unhealthy individuals with a high demand for doctor visits have a higher probability to drop out of the data set due to severe illness, death, or other reasons. This is a problem because it can be expected that the statutory health insurance covers a higher share of unhealthy people compared to the private insurance because of the worse risk-pool. We follow Freund et al. (1999) and include an inverse Mills ratio to control for possible panel attrition. A natural problem with panel attrition is that no individual characteristics can be observed when a person has already left the panel. Only the information of not being in the panel can be observed for attritors. Therefore, the inverse Mills ratio is constructed using the estimates of a probit regression of an indicator to be in the panel in the next period on all current control variables that appear later in the regression model. Variables that are assumed to have an effect on panel attrition but not on the number of doctor visits (conditional on the other control variables) are also included here, namely the degree of life satisfaction, a dummy for oral interview (instead of written), and the duration of the interview. Life satisfaction should not affect the demand for doctor visits once health satisfaction is controlled for but is likely to affect the general likelihood of participating in a panel study.⁸ Likewise, if the interview was oral and shorter, the probability of staying in the panel is assumed to increase, whereas these characteristics should not affect the number of physician visits.

⁸See Frijters et al. (2004) who find panel attrition in the SOEP that is negatively linked to life satisfaction.

5 Estimation Results

Following the Akaike information criterion (AIC), a mixture model with four mass points outperformed models with three and two mass points, and the basic model without control for unobserved heterogeneity. The main results, however, are similar to all other models with a smaller number of mass points. A model with five mass points failed to converge. The regression results are reported in Table 3.

Due to the non-linearity of the model, the estimated coefficients cannot be interpreted as marginal effects. However, since the underlying models are a logit and a negative binomial model, signs and significance can be directly interpreted. Looking at the differences between publicly and privately insured over time (i.e., looking at the coefficients of $SHI \times T_{93-97}$, $SHI \times T_{98-99}$, and $SHI \times T_{00-03}$), we find that the basic effects do not differ strongly between males and females. The coefficients in the first stage are always insignificant for both men and women. The effect of the introduction of a fixed budget in 1993 is also insignificant in the second stage, thus confirming the expected result of no effects of this reform on the number of doctor visits. However, there are highly significant negative coefficients of the 1997 reform in the second stage. The period of increased tightness of the national budget (after 2000) is again associated with a negative coefficient for the publicly insured which, however, is not statistically different from the coefficients of the 1997 reform. Therefore, these results provide some evidence that the effect of the 1997 reform was not only a short-term one.

To get a better idea of the effects, we use the regression results and look at the time trends of the number of doctor visits when compositional differences and macroeconomic conditions are controlled for. Figures 3 and 4 show predicted values for the first and the second stages, holding all characteristics fixed at the sample averages and varying only the time and the insurance group indicators. Furthermore, the predicted values are averaged over the four latent types, weighted by their respective probabilities. That is, we compare two hypothetical average individuals that only differ in their insurance status (and not, for instance, in their health status). No real variation over time can be found in the first stage (Figures 3a and 4a) for males or for females.⁹ Conditional on observed characteristics like the health status, the publicly insured have a slightly higher likelihood of one visit throughout the entire observation period. This might reflect incentive effects of the insured due to the absence of co-payments in the public system.

However, there are remarkable evolutions in the second stage. Note that Figures 3b and 4b do not show the clear decreasing time trend that was found in the un-

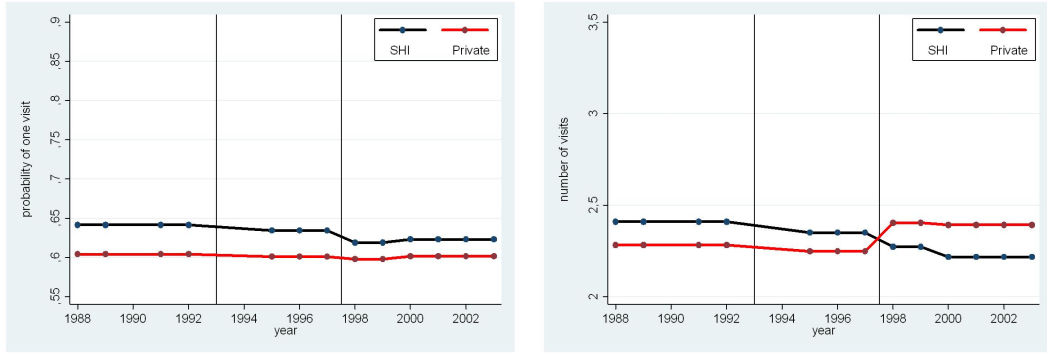
⁹Note that the slight drop of the SHI-group after 1997 in Figure 3a is not significant.

Table 3: Estimated coefficients of the finite mixture hurdle model

	Males				Females			
	Stage 1		Stage 2		Stage 1		Stage 2	
$SHI \times T_{93-97}$	-0,022	(0,10)	-0,015	(0,06)	-0,031	(0,13)	-0,057	(0,06)
$SHI \times T_{98-99}$	-0,085	(0,10)	-0,192*	(0,06)	-0,043	(0,14)	-0,158*	(0,06)
$SHI \times T_{00-03}$	-0,082	(0,08)	-0,229*	(0,05)	0,118	(0,11)	-0,175*	(0,05)
T_{93-97}	-0,015	(0,10)	-0,028	(0,06)	0,103	(0,13)	0,043	(0,06)
T_{98-99}	-0,032	(0,10)	0,091	(0,06)	0,033	(0,14)	0,095	(0,06)
T_{00-03}	-0,014	(0,08)	0,082	(0,05)	-0,164	(0,11)	0,071	(0,05)
SHI	0,193*	(0,07)	0,095*	(0,04)	0,067	(0,10)	0,113*	(0,04)
Health Satisfaction	-0,269*	(0,01)	-0,166*	(0,00)	-0,252*	(0,01)	-0,143*	(0,00)
Disabled	0,855*	(0,04)	0,302*	(0,02)	0,787*	(0,06)	0,290*	(0,01)
Hospital visit previous year	0,049*	(0,01)	0,010*	(0,00)	0,038*	(0,01)	0,012*	(0,00)
Sport	0,094*	(0,01)	0,006	(0,01)	0,076*	(0,01)	0,003	(0,00)
Unemployment rate	-0,004	(0,01)	0,008	(0,01)	-0,025*	(0,01)	0,002	(0,00)
Absent > 3 days	0,847*	(0,02)	0,329*	(0,01)	0,745*	(0,03)	0,300*	(0,01)
Age	-0,044*	(0,01)	0,009*	(0,00)	-0,048*	(0,01)	0,005	(0,00)
Age Squared	0,001*	(0,00)	-0,000	(0,00)	0,001*	(0,00)	-0,000	(0,00)
Married	0,209*	(0,03)	0,018	(0,02)	0,173*	(0,03)	0,061*	(0,01)
Children under 16	-0,151*	(0,03)	-0,060*	(0,02)	-0,083*	(0,03)	-0,054*	(0,01)
Small town	-0,067*	(0,03)	-0,038*	(0,02)	-0,038	(0,03)	-0,037*	(0,01)
Large town	0,011	(0,03)	0,061*	(0,02)	0,060	(0,03)	0,032*	(0,01)
Years of schooling	0,035*	(0,01)	-0,008*	(0,00)	0,042*	(0,01)	0,007*	(0,00)
Full-time employed	-0,159*	(0,06)	-0,081*	(0,04)	-0,304*	(0,05)	-0,153*	(0,02)
Part-time employed	-0,036	(0,10)	0,058	(0,05)	-0,246*	(0,05)	-0,121*	(0,02)
Unemployed	-0,304*	(0,05)	0,021	(0,02)	-0,205*	(0,04)	-0,008	(0,02)
Blue collar worker	-0,380*	(0,06)	-0,151*	(0,04)	-0,267*	(0,05)	-0,105*	(0,02)
White collar worker	-0,262*	(0,06)	-0,189*	(0,04)	-0,023	(0,05)	-0,121*	(0,02)
Self-employed	-0,513*	(0,07)	-0,259*	(0,04)	-0,322*	(0,07)	-0,157*	(0,03)
Health job	-0,135	(0,11)	-0,143	(0,08)	-0,177*	(0,05)	-0,037	(0,03)
Civil	-0,055	(0,08)	-0,090	(0,05)	0,134	(0,10)	-0,087*	(0,04)
Net-household inc./1000	0,011	(0,01)	-0,004	(0,00)	-0,003	(0,01)	-0,014*	(0,00)
West Germany	-0,163	(0,09)	0,215*	(0,05)	-0,358*	(0,09)	0,101*	(0,04)
Inverse Mills ratio	-0,670	(0,36)	0,684*	(0,20)	-1,198*	(0,43)	0,650*	(0,17)
α_1			5.062*	(0.672)			0.827*	(0.023)
α_2			0.418*	(0.015)			0.091*	(0.012)
α_3			0.213*	(0.012)			0.285*	(0.010)
α_4			10.208*	(0.462)			13.002*	(0.520)
θ^1	0.523*		(0.092)		1.069*		(0.019)	
θ^2	0.990*		(0.022)		-0.047*		(0.019)	
θ^3	0.246*		(0.020)		0.531*		(0.016)	
θ^4	- 1.366*		(0.018)		1.131*		(0.016)	
$P(\theta^1, \alpha_1)$	0.095*		(0.01)		0.145*		(0.01)	
$P(\theta^2, \alpha_2)$	0.227*		(0.01)		0.267*		(0.01)	
$P(\theta^3, \alpha_3)$	0.386*		(0.01)		0.367*		(0.01)	
$P(\theta^4, \alpha_4)$	0.291*		(0.01)		0.221*		(0.01)	
Log-pseudolikelihood	-147254.915				-182720.336			
Akaike	294657.830				365588.672			
Observations	82,621				88,575			

* indicates significance at the 5% level

Figure 3: Predicted values males

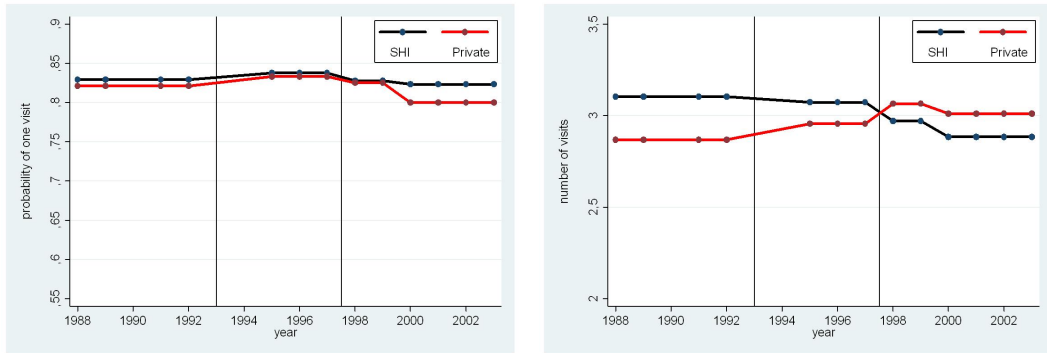


(a) 1st stage

(b) 2nd stage

Predicted values for average individuals in the sample, based on regression results in Table 3.

Figure 4: Predicted values females



(a) 1st stage

(b) 2nd stage

Predicted values for average individuals in the sample, based on regression results in Table 3.

conditional numbers in Figures 1 and 2. Until 1997, the publicly insured had more conditional doctor visits than comparable privately insured. Not only did their number of visits decrease after 1997, the number of visits in the privately insured group also increased. Since there is no reason to assume that there was an unobserved shock that affected only the privately insured and shifted their number of visits to a higher level, this evolution seems to be a result of a change in the patient mix by the physicians. This important identifying assumption is also supported by the evolution after 1997. While before the reform the difference between the privately and the publicly insured is stable (more visits of the publicly insured in the second stage), it is also stable afterwards (more visits of the privately insured). The change after 1997 is a long-term change which is unlikely to result from time-varying unobserved effects. Looking at these figures, the assumption of no general equilibrium effects becomes very unlikely to hold true.

In a series of articles, Winkelmann (2004a, 2004b, 2006) also uses the SOEP to estimate the effect of increased co-payments for prescription drugs on the number of

doctor visits. In July 1997, the co-payments for prescription drugs were increased for the publicly insured by 6 DM (about 3 Euro) which, depending on the package size was an increase of about 86% (from 7 DM to 13 DM for large sizes) to 200% (from 3 DM to 9 DM for small sizes). The privately insured were not affected by the reform. Winkelmann uses them as a control group and finds a reduction in the expected number of doctor visits by about 10% between 1995 and 1999. In order to make our results comparable to the ones of Winkelmann, we use the results of our regression and calculate a “treatment effect”, the derivation of which is explained in the appendix.¹⁰

The first column of Table 4 reports the average treatment effect on the treated. Although it is only unbiased if one is willing to assume that there are no general equilibrium effects, we report it as a benchmark to compare it to Winkelmann’s results. Signs, significance, and relative sizes of the coefficients to each other do not differ to the results in Table 3. However, here, the sizes can be directly quantified. The estimated treatment effect for the reform of 1993 is zero. The one for the reform of 1997, however, is strong and significant. The values of about 0.22 for males and 0.32 for females mean an approximate drop of 10% in the number of doctor visits due to the reform. Thus, assuming that the privately insured are a proper control group, we can replicate Winkelmann’s finding. The question is, whether this effect results from the remuneration reform or the co-payment reform.

Our hurdle model specification allows us not only to calculate the overall effect but to decompose it into parts that are due to changes in the first stage and those due to changes in the second stage. The third and the fifth columns in Table 4 report these effects. The effects in the first stage are very small and insignificant for all three reforms. Those in the second stage for the reform of 1997 and the time thereafter, however, are quite large and statistically significant. Thus, the negative overall treatment effect is exclusively the result of a negative effect in the second stage. The motivation of the hurdle model as a model with two different decision-making processes – the patient has full control in the first stage but the physician takes over in the second stage – supports the interpretation that the reform did not affect the patient’s but only the physician’s behaviour.

Winkelmann argues that the co-payment reform might have lowered the probability of doctor visits, because individuals either do not go to the doctor anymore because they fear getting a prescription, or that they have less visits than before, because they demand prescriptions for larger package sizes. The first argument

¹⁰We are aware of the problem with this terminology. Although it is not clear, what this quantity really measures in this case, we call this a “treatment effect”. We do this to keep things simple because our specification is similar to a difference-in-differences estimation.

Table 4: Estimated Treatment Effects

		Overall	s.e.	1st stage	s.e.	2nd stage	s.e.
Males	Reform 1993	-0.024	(0.071)	-0.004	(0.019)	-0.021	(0.082)
	Reform 1997	-0.221*	(0.084)	-0.016	(0.020)	-0.270*	(0.098)
	After 2000	-0.249*	(0.067)	-0.016	(0.015)	-0.313*	(0.078)
Females	Reform 1993	-0.123	(0.120)	-0.004	(0.015)	-0.129	(0.128)
	Reform 1997	-0.322*	(0.141)	-0.005	(0.017)	-0.355*	(0.150)
	After 2000	-0.284*	(0.108)	0.015	(0.014)	-0.381*	(0.114)

Treatment effects calculated according to equation (2) in the appendix based on regression results in Table 3. All control variables (except for the interesting reform dummies) set to represent the average individual in the data set. * indicates significance at the 5% level

seems disputable since this is a very indirect effect and it is hard to imagine that this would reduce the total number of doctor visits by such a great amount. While the co-payment increase is high in relative terms, it is rather low (about 3 Euro) in absolute terms. Other studies show that even more direct and somewhat stronger demand side incentives have much smaller effects. In an evaluation of the introduction of a 10 Euro co-payment for doctor visits, Farbmacher (2009) finds a reduction in the probability of one doctor visit (i.e., in the first stage) of up to 3.4% while there are no effects in the positive part of the distribution (the second stage). Moreover, apart from the co-payments for doctor visits, also in 2004, there was another increase in co-payments for prescription drugs. By attributing the entire reduction to the co-payments for doctor visits, Farbmacher (2009) implicitly finds no effects of increased co-payments for prescription drugs here. The second of Winkelmann's arguments (patients demand higher package sizes to reduce the number of doctor visits) is more likely to be true. However, because co-payments for large package sizes are still about 50% higher than for small sizes, the incentives to switch from smaller to larger package sizes might not be very strong. All in all, the incentives imposed by the prescription-drug reform are quite low and such a strong reaction – a 10% drop in the number of doctor visits – seems to be surprising.

More importantly, the increased co-payments should lead to a demand-side effect only. Turning back to Table 4 and Figures 3 and 4, the results for the first stage are practically zero. The decreasing expected number of doctor visits only result from the changes in the second stage.¹¹ Even though it is possible that the patient

¹¹Winkelmann (2004b) also uses a two-part model to distinguish the effects on the first and the second stage. Contrary to our study, he finds significant effects in both stages, and even a bigger one in the first stage. His study, however, does not use a control group but is just a before-after comparison. Without a control group, however, one has to impose the rather strict assumption of absence of exogenous temporary shocks that affect the number of doctor visits. Turning back to Figure 2a, we can see a drop in the probability of one doctor visit between 1996 and 1998. This, however, applies to both statutorily and privately insured. Winkelmann (2004a) uses a control group, but does not present results of the hurdle model, hence one cannot separate a demand from a supply-side effect. Winkelmann (2006) uses the method of quantile regressions for count data and

also has some level of control in the second stage apart from the physician, the findings from the first stage indicate that it is unlikely that the second-stage results are driven by demand-side effects.

Having said this, looking back at the evolution in Figures 3 and 4, we directly see that the estimated “treatment effect” of a decrease in the number of doctor visits by 10% is upward biased. This strong result does not only follow from a drop in the visits of the publicly insured but to a great deal also from an increase in the visits of the “control group”. Therefore, no difference-in-differences analysis is possible here when the effect of reforms on the *absolute* number of doctor visits is to be evaluated. However, we can say that physicians changed their patient-mix. While until 1997, the publicly insured had more conditional doctor visits than the comparable privately insured, this picture turned around immediately after the reform took effect.

Other Covariates

Not surprisingly, given the high number of observations, most of the other control variables are highly significant (Table 3). Here, we only discuss the most interesting ones. The variables that capture the observed health status (health satisfaction, hospital stays in the last year, at least three days of absence, handicap) are all highly significant in both stages and have the expected signs. Individuals who frequently do sports have a higher probability of visiting a doctor. Conditional on the health status this can be interpreted as a higher concern for own health of these individuals. This interpretation is supported by the finding that *sport* is not significant in the second stage. West Germans have a lower probability of visiting a doctor than East Germans. This remarkable difference can be interpreted by preventive doctor visits that used to have a much higher importance in the former GDR and still have in East Germany. However, the conditional number of visits is much higher in West Germany.¹² Individuals who live in large towns have more doctors visits. This might be explained by better access to services in larger towns (should affect the first stage) but also increased competition between doctors in urban regions,

the years 1996 and 1998 only. Using a control group here, he finds that the reform effect was larger in the lower part of the distribution (the “low-users”) than in the upper part (the “high-users”). This strategy allows to analyse different behavioural responses of heterogenous individuals. However, arguably, the two-part model might be better suited to separate demand from supply-side effects than a quantile regression model.

¹²This explains the drop in doctor visits after 1991 in Figures 1 and 2b. This is a compositional effect because before 1992 there are only West Germans in the sample. Note the absence of this drop in Figures 3 and 4 after conditioning on the region.

resulting in a higher degree of physician-induced demand here (second stage). The unemployment rate also shows the expected sign, although it is only significant for females. A higher unemployment rate is associated with a lower number of doctor visits (this effect is partly also captured by the absenteeism indicator which, however, also captures the individual health status). Note, however, that it only affects the first stage, thus being in line with the interpretation of individuals deciding not to see a doctor in times of recessions.

6 Robustness Checks

6.1 Subsamples and Specifications

The identification of the reform effect on the patient-mix rests on some assumptions. The first one is that the trends in doctor visiting behaviour would have stayed stable without the reform in 1997. This cannot be tested directly but a test used by Galiani et al. (2005) could give a notion of how credible this assumption is. Basically, this test consists of using only the pre-reform years, running placebo-reforms and testing whether one finds differences between control and treatment groups before. Since we already included an interaction-dummy for the period of 1993 to 1997 in the regression ($SHI \times T_{93-97}$) which is not significant, we directly infer that the assumption of parallel trends is supported.

As discussed earlier, switching insurance from public to private or vice versa is not possible for the majority of individuals. Moreover, switching to private insurance as a reaction to the reforms seems unlikely because it is especially the sicker individuals who would be punished by paying a higher insurance premium in the private sector. Nevertheless, as a second robustness check, to completely rule out endogeneity problems of the treatment, we dropped all individuals from the sample who switched at least once their insurance status after 1996. The results for these two subsamples are reported in Table 5 in Columns 1 and 2. Neither in the first nor in the second stage do we see important differences in the estimated effects.

We experimented with several other specifications and subsamples to check the robustness of the results. As pointed out in the text, apart from a baseline model without a random effect, we also estimated models with two and three mass points. The model with four mass points is preferred for statistical reasons but does not lead to results different to those of the other models. As shown by Bertrand et al. (2004), difference-in-differences estimations tend to underestimate the standard errors of the treatment effect as the included number of years after the treatment increases. This does not seem to be a problem here since for the most interesting treatment effect (the reform of 1997), only two post-reform years are used and

Table 5: Non-changers

	Males		Females	
	Stage 1	Stage 2	Stage 1	Stage 2
$SHI \times T_{93-97}$	-0.010 (0.11)	-0.030 (0.07)	0.022 (0.15)	-0.004 (0.07)
$SHI \times T_{98-99}$	-0.055 (0.12)	-0.283* (0.07)	0.095 (0.17)	-0.167* (0.08)
$SHI \times T_{00-03}$	-0.046 (0.09)	-0.276* (0.05)	0.084 (0.13)	-0.201* (0.06)
T_{93-97}	-0.045 (0.11)	-0.015 (0.07)	0.034 (0.15)	-0.012 (0.07)
T_{98-99}	-0.065 (0.12)	0.182* (0.07)	-0.117 (0.18)	0.107 (0.08)
T_{00-03}	-0.056 (0.09)	0.132* (0.05)	-0.141 (0.13)	0.098 (0.06)
SHI	0.161 (0.08)	0.113* (0.05)	0.012 (0.12)	0.111* (0.05)
Observations	74,942		82,728	

Asterisk indicates significance at the 5% level. Coefficients of all other covariates not shown here.

moreover, the estimated z-statistics are rather high. Nevertheless, we re-estimated the same model using only the years 1996 (pre-reform) and 1998 (post-reform).¹³ The results do not change qualitatively (i.e., no significant effect in stage 1 but a significant reduction in stage 2).

6.2 Test for Single Spell Assumption

As mentioned above, the analysis using the hurdle model rests on the assumption that the observed number of doctor visits in a quarter results from one sickness spell only. This “single spell assumption” can be verified by a test which was derived by Santos Silva and Windmeijer (2001). We only briefly outline the procedure here and refer to Santos Silva and Windmeijer (2001) for a more detailed derivation and description. The notation here also follows Santos Silva and Windmeijer (2001). Let V be the number of doctor visits in the previous three months. If S denotes the number of illness spells and R_j the number of recalls in spell j , V can be expressed as $V = \sum_{j=1}^S R_j = S + \sum_{j=1}^S R_j^*$, where $R_j^* = R_j - 1$. Assuming that S and R_j are conditionally independent and with $E(R_j|x, \gamma) = E(R|x, \gamma)$, $j = 1, \dots, S$, one can show that

$$E(V|x, \beta, \gamma) = E(S|x, \beta)E(R|x, \gamma).$$

¹³Estimation results of the robustness checks are not presented here but available upon request.

We only observe V but want to identify the parameters β and γ of $E(S|x, \beta)$ and $E(R|x, \gamma)$. This is only possible if S does not exceed 1, that is, if $S = d = \min\{V, 1\}$. If this is true, the following moment condition holds:

$$E[d - E(S|x, \beta)] = 0. \quad (1)$$

This condition can be used to estimate β . Likewise, since for the positive counts

$$E(V|V > 0, S \leq 1, x, \gamma) = E(R|x, \gamma),$$

the following moment condition can be used to estimate γ :

$$E\{[V - E(R|x, \gamma)]|V > 0\} = 0. \quad (2)$$

The idea of the test for the single spell hypothesis is to estimate β and γ by GMM using the moment conditions (1) and (2) and specification of the first moments and then to test whether the following expression significantly differs from zero when the estimated parameters are plugged in, i.e., to test if the following holds:

$$E[m(V, x, \beta, \gamma)] = E[V - E(S|x, \beta)E(R|x, \gamma)] = 0.$$

This type of conditional moments test is described in, e.g., Cameron and Trivedi (2005). Under the null hypothesis of $E[m(V, x, \beta, \gamma)] = 0$, the test statistic M follows a chi-square distribution with h degrees of freedom where h is the number of moment conditions, i.e., the number of included variables in x :

$$M = \hat{m}(\hat{\beta}, \hat{\gamma})' [\hat{V}\{\hat{m}(\hat{\beta}, \hat{\gamma})\}]^{-1} \hat{m}(\hat{\beta}, \hat{\gamma}) \sim \chi^2(h)$$

Here, in order to stay consistent with the negative binomial model in section 4.1, we specify

$$E(S|x, \beta) = 1 - \exp(-\exp(x'\beta))$$

and

$$E(R|x, \gamma) = \frac{\exp(x'\gamma)}{1 - \exp(-\exp(x'\gamma))}.$$

We bootstrap the estimate of the variance-covariance matrix of m with 500 replications. A test statistic above the critical value leads to a rejection of the null. Consequently, failure to reject the null hypothesis would then be an evidence in favour of the single spell assumption. Since the test by Santos Silva and Windmeijer (2001) is derived for cross-sections, we carry out separate tests for every year.

We include the same covariates as in the regression analysis but have to leave out all those which do not vary between individuals within a given year (essentially, these are the time and treatment effect dummies).

Table 6 reports the results of the single spell test. The test statistic is below the critical value of the chi-square distribution with a significance level of 5% for all cases. Thus, we cannot reject the single spell hypothesis.

Table 6: Test for single spell assumption

Year	Test statistic		Critical value
	Males	Females	
1988	24.909	12.899	36.415
1989	21.562	28.765	36.415
1991	24.737	28.796	36.415
1992	29.391	36.639	37.652
1995	25.026	32.811	37.652
1996	33.872	22.739	37.652
1997	27.474	30.590	37.652
1998	25.635	36.362	37.652
1999	36.069	23.129	37.652
2000	15.322	34.767	37.652
2001	27.273	33.080	37.652
2002	30.449	29.758	37.652
2003	27.084	31.638	37.652

Critical values are of chi-square distribution with 24 or 25 degrees of freedom and 5% significance level. 25 degrees of freedom after 1991 because a dummy for West Germany is included only thereafter.

To our knowledge, we are the first to apply this test apart from Santos Silva and Windmeijer (2001) who used it with data from 1985 of the SOEP.¹⁴ Therefore, no evidence from other data sets is available with which to compare this result. However, we think that the short period of three months within which the number of visits to the doctor is measured leads to this finding. While in most other household panel data sets, the period is one year, three months seem to be short enough to justify the single spell assumption. Based on this result, we argue that differentiation between first and second stage, which is essential for our identification strategy, is possible given the data at hand.

7 Conclusion

Two major reforms affected the supply-side of ambulatory care in Germany in the last 15 years, namely the introduction of a fixed budget in 1993 and of individual practice budgets for physicians in 1997. With data from the German Socio-Economic

¹⁴The authors reject the assumption for specialist visits but fail to reject it for GPs.

Panel that cover the period 1988-2003, we find no effect of the introduction of a fixed budget but a strong effect of the individual practice budgets on the number of doctor visits of both publicly and privately insured.

The results show that the behaviour of patients has not changed due to the reforms since the likelihood of one visit to the doctor within a period of three months remained stable. However, the number of recalls changed gradually. While until 1997, publicly insured patients had more doctor visits than comparable privately insured individuals, given they had seen a doctor at least once, this picture turned around after the second reform became effective. After the reform, privately insured patients have more visits when characteristics like the health status are controlled for. The results hold for both males and females and are robust to several subsamples.

The results are in line with a general notion in the German public of the privately insured getting preferential treatment from physicians at the cost of the publicly insured (see Lungen et al. (2008) for an experimental study on the access to physicians of the privately and publicly insured). This study gives evidence on how the remuneration system can directly affect physician behaviour. Since we control for patient characteristics like the health status, education or income, the strong variation in the number of recalls displays the physicians' influence in controlling the demand for treatments. Therefore, this finding is also a hint at the existence of supplier-induced demand. Since we did not evaluate health changes due to the reforms, we cannot say if physicians reduced excess and supposedly trivial treatments in the group of publicly insured after 1997 or if they also reduced the number of necessary treatments. The increased number of visits among privately insured which cannot be attributed to a shock like a sudden drop of health status, however, points to an increased number of unnecessary treatments in this group. A full evaluation of the health effects of these reforms is left for future research.

Previous German studies only find weak or no effects of demand-side incentives on cost-saving behaviour of the insured (see, e.g., Pohlmeier and Ulrich, 1995, Riphahn et al., 2003, Augurzky et al., 2006, Schreyögg and Grabka, 2008, Farbmacher, 2009; however also see Felder and Werblow, 2008). The findings of the current study suggest that doctors react strongly to incentives. Therefore, in general, supply-side regulations might be a much better instrument to manage health care expenditures than adjusting incentives on the demand side in Germany. Using the physician remuneration system to reduce the inefficiency implied by the information asymmetry between doctor and patient seems to be a good starting point in order to maintain universal access to health care at a reasonable cost.

References

- Askildsen, Jan Erik, Espen Bratberg, and Øivind Anti Nilsen, “Unemployment, labor force composition and sickness absence: a panel data study,” *Health Economics*, 2005, *14* (11), 1087–1101.
- Augurzky, Boris, Thomas K. Bauer, and Sandra Schaffner, “Copayments in the German health system: does it work?,” *IZA Discussion Paper*, September 2006, (2290).
- Bago d’Uva, Teresa, “Latent class models for health care utilisation,” *Health Economics*, 2006, *15*, 329–343.
- and Andrew M. Jones, “Health care utilisation in Europe: New evidence from the ECHP,” *Journal of Health Economics*, March 2009, *28* (2), 265–279.
- Benstetter, Franz and Achim Wambach, “The treadmill effect in a fixed budget system,” *Journal of Health Economics*, January 2006, *25* (1), 146–169.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, “How much should we trust differences-in-differences estimates?,” *Quarterly Journal of Economics*, February 2004, *119* (1), 249–275.
- Breyer, Friedrich, “Die Nachfrage nach medizinischen Leistungen: Eine empirische Analyse von Daten aus der Gesetzlichen Krankenversicherung,” *Berlin*, 1984.
- Cameron, A. Colin and Pravin K. Trivedi, *Microeconometrics: Methods and Applications*, New York, NY, USA: Cambridge University Press, 2005.
- Contoyannis, Paul, Andrew M. Jones, and Nigel Rice, “The dynamics of health in the British Household Panel Survey,” *Journal of Applied Econometrics*, 2004, *19* (4), 473–503.
- Croxson, B., C. Propper, and A. Perkins, “Do doctors respond to financial incentives? UK family doctors and the GP fundholder scheme,” *Journal of Public Economics*, February 2001, *79* (2), 375–398.
- Deb, Partha and Pravin K. Trivedi, “Demand for Medical Care by the Elderly: A Finite Mixture Approach,” *Journal of Applied Econometrics*, May-June 1997, *12* (3), 313–36.
- and – , “The structure of demand for health care: latent class versus two-part models,” *Journal of Health Economics*, July 2002, *21* (4), 601–625.
- Devlin, Rose Anne and Sisira Sarma, “Do physician remuneration schemes matter? The case of Canadian family physicians,” *Journal of Health Economics*, September 2008, *27* (5), 1168–1181.
- Dusheiko, Mark, Hugh Gravelle, Rowena Jacobs, and Peter Smith, “The effect of financial incentives on gatekeeping doctors: Evidence from a natural experiment,” *Journal of Health Economics*, May 2006, *25* (3), 449–478.

- Farbmacher, Helmut**, “Copayments for doctor visits in Germany and the probability of visiting a physician - Evidence from a natural experiment,” *Discussion paper*, July 2009, (2009-10).
- Felder, Stefan and Andreas Werblow**, “A Physician Fee that Applies to Acute but not to Preventive Care: Evidence from a German Deductible Program,” *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 2008, 128 (2), 191–212.
- Freund, Deborah A., Thomas J. Kniesner, and Anthony T. LoSasso**, “Dealing with the common econometric problems of count data with excess zeros, endogenous treatment effects, and attrition bias,” *Economics Letters*, January 1999, 62 (1), 7–12.
- Frijters, Paul, John P. Haisken-DeNew, and Michael A. Shields**, “Money does matter! Evidence from increasing real income and life satisfaction in East Germany following reunification,” *American Economic Review*, June 2004, 94 (3), 730–740.
- , – , and – , “The causal effect of income on health: Evidence from German reunification,” *Journal of Health Economics*, September 2005, 24 (5), 997–1017.
- Galiani, Sebastian, Paul Gertler, and Ernesto Schargrotsky**, “Water for life: the impact of the privatization of water services on child mortality,” *Journal of Political Economy*, February 2005, 113 (1), 83–120.
- Grabka, Markus M.**, “Prämien in der PKV : deutlich stärkerer Anstieg als in der gesetzlichen Krankenversicherung,” *Wochenbericht*, 2006, 73 (46), 653–659.
- Grytten, Jostein and Rune Sorensen**, “Type of contract and supplier-induced demand for primary physicians in Norway,” *Journal of Health Economics*, May 2001, 20 (3), 379–393.
- Haisken-DeNew, John P. and Markus Hahn**, “PanelWhiz: A flexible modularized Stata interface for accessing large-scale panel data sets,” 2006.
- Hennig-Schmidt, Heike, Reinhard Selten, and Daniel Wiesen**, “Physicians’ Provision Behavior under Different Payment Systems - An Experimental Investigation,” http://www.uni-bonn.de/~wiesen/physician_behavior.pdf, 2008.
- Jimenez-Martin, Sergi, Jose M. Labeaga, and Maite Martinez-Granado**, “Latent class versus two-part models in the demand for physician services across the European Union,” *Health Economics*, 2002, 11 (4), 301–321.
- Jones, A., N. Rice, T. Bago d’Uva, and S. Balia**, *Applied Health Economics* 2007.
- Jürges, Hendrik**, “Health insurance status and physician behavior in Germany,” *Journal of Applied Social Science Studies (Schmollers Jahrbuch)*, 2009, 129 (2).

- Kopetsch, Thomas**, “Arztdichte und Inanspruchnahme ärztlicher Leistungen in Deutschland: Eine empirische Untersuchung der These von der angebotsinduzierten Nachfrage nach ambulanten Arztleistungen,” *Schmollers Jahrbuch*, 2007, 127, 373–405.
- Krämer, Walter**, “Eine ökonometrische Untersuchung des Marktes für ambulante kassenärztliche Leistungen,” *Zeitschrift für die gesamte Staatswissenschaft*, 1981, 1, 45–61.
- Lungen, Markus, Bjoern Stollenwerk, Philipp Messner, Karl W Lauterbach, and Andreas Gerber**, “Waiting times for elective treatments according to insurance status: A randomized empirical study in Germany,” *International Journal for Equity in Health*, January 2008, 7 (1), 1–7.
- Madden, David, Anne Nolan, and Brian Nolan**, “GP reimbursement and visiting behaviour in Ireland,” *Health Economics*, 2005, 14 (10), 1047–1060.
- Mullahy, John**, “Specification and testing of some modified count data models,” *Journal of Econometrics*, December 1986, 33 (3), 341–365.
- OECD**, “Health Data 2008 - Version: June 2008,” 2008.
- Pohlmeier, Winfried and Volker Ulrich**, “An econometric model of the two-part decision-making process in the demand for medical care,” *Journal of Human Resources*, 1995, 30, 339–361.
- Puhani, Patrick A.**, “The treatment effect, the cross difference, and the interaction term in nonlinear ‘difference-in-differences’ models,” *IZA Discussion Paper*, April 2008, (3478).
- Riphahn, Regina T., Achim Wambach, and Andreas Million**, “Incentive effects in the demand for health care: a bivariate panel count data estimation,” *Journal of Applied Econometrics*, 2003, 18 (4), 387–405.
- Santos Silva, Joao M. C. and Frank Windmeijer**, “Two-part multiple spell models for health care demand,” *Journal of Econometrics*, August 2001, 104 (1), 67–89.
- Schreyögg, Jonas and Markus M. Grabka**, “Copayments for ambulatory care in Germany: a natural experiment using a difference-in-differences approach,” *SOEPpapers*, 2008, (96).
- Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp**, “The German Socio-Economic Panel Study (SOEP): Scope, evolution, and enhancements,” *Journal of Applied Social Science (Schmollers Jahrbuch)*, 2007, 127 (1), 139–169.
- Winkelmann, Rainer**, “Co-payments for prescription drugs and the demand for doctor visits—evidence from a natural experiment,” *Health Economics*, 2004, 13 (11), 1081–1089.
- , “Health care reform and the number of doctor visits—an econometric analysis,” *Journal of Applied Econometrics*, 2004, 19 (4), 455–472.

– , “Reforming health care: Evidence from quantile regressions for counts,” *Journal of Health Economics*, January 2006, *25* (1), 131–145.

Wittek, Lothar, “Die Mengenentwicklung hat viele Ursachen,” *Deutsches Ärzteblatt*, June 1996, *93* (24).

Wörz, Markus and Reinhard Busse, “Analysing the impact of health-care system change in the EU member states - Germany,” *Health Economics*, 2005, *14* (S1), S133–S149.

Appendix

Derivation of the Treatment Effect with Nonlinear Difference-in-Differences

Let Y be the observed outcome variable (number of doctor visits), G an indicator of the treatment group, and T a dummy variable for the post-treatment time.¹⁵ Then the treatment effect on the treated is defined as

$$\tau(T = 1, G = 1, X) = E[Y^1|T = 1, G = 1, X] - E[Y^0|T = 1, G = 1, X]$$

where Y^1 is the observed outcome and Y^0 is the potential outcome in the absence of the treatment (the contrafactual situation). With a control group and the identifying assumption that the differences between treatment and control group would have been stable over time, this expression reduces to the parameter of the interaction term $T \times G$ in a linear regression of Y on T , G , $T \times G$, and other control variables. This, however, is not possible in a nonlinear model. Let the conditional expectation of Y depend on a possibly nonlinear function g

$$E[Y|T, G, X] = g(\alpha T + \beta G + \gamma TG + X\theta)$$

Then the treatment effect is given as (see Puhani, 2008)

$$\tau(T = 1, G = 1, X) = g(\alpha + \beta + \gamma + X\theta) - g(\alpha + \beta + X\theta)$$

The expectation in the hurdle specification is given by the probability of a positive count (1st stage; call this function ψ_1) times the expectation conditional on positive counts (2nd stage; call this function ψ_2)¹⁶:

$$\begin{aligned} E[Y|T, G, X] &= P(Y > 0|T, G, X) * E[Y|Y > 0, T, G, X] \\ &= \psi_1(\alpha_1 T + \beta_1 G + \gamma_1 TG + X\theta_1) * \psi_2(\alpha_2 T + \beta_2 G + \gamma_2 TG + X\theta_2) \end{aligned} \quad (1)$$

Then, the treatment effect on the treated, together with the hurdle specification, is given by

$$\begin{aligned} \tau(T = 1, G = 1, X) &= \psi_1(\alpha_1 + \beta_1 + \gamma_1 + X\theta_1) * \psi_2(\alpha_2 + \beta_2 + \gamma_2 + X\theta_2) \\ &\quad - \psi_1(\alpha_1 + \beta_1 + X\theta_1) * \psi_2(\alpha_2 + \beta_2 + X\theta_2) \end{aligned} \quad (2)$$

Unlike in the linear case, the treatment effect is not constant over all individuals.

¹⁵The notation and the general notion of nonlinear difference-in-differences closely follow Puhani (2008).

¹⁶Note that both functions in the hurdle model are strictly monotonic transformations of the linear index and, thus, also the product of the two.

Here, it is calculated according to (2), where all the control variables (except for the interesting reform dummies) are set to represent the average individual in the dataset.

Table A1: Probability of staying in the sample

	Coefficient	Std. Error	Coefficient	Std. Error
$SHI \times T_{93-97}$	-0,106	(0,067)	-0,116	(0,073)
$SHI \times T_{98-99}$	0,057	(0,069)	-0,297*	(0,089)
$SHI \times T_{00-03}$	-0,011	(0,050)	-0,281*	(0,058)
T_{93-97}	0,188*	(0,065)	0,163*	(0,072)
T_{98-99}	0,108	(0,067)	0,457*	(0,088)
T_{00-03}	0,169*	(0,049)	0,432*	(0,057)
SHI	0,022	(0,046)	0,359*	(0,050)
Health Satisfaction	0,016*	(0,004)	0,021*	(0,003)
Disabled	0,009	(0,021)	0,074*	(0,024)
Hospital Stays prev. year	-0,004*	(0,002)	-0,006*	(0,002)
Sport	0,010	(0,006)	0,005	(0,006)
Unemployment rate	0,021*	(0,006)	0,022*	(0,006)
Absent > 3 days	0,046*	(0,015)	0,046*	(0,017)
Age	0,040*	(0,003)	0,043*	(0,002)
Age Squared	-0,000*	(0,000)	-0,000*	(0,000)
Married	0,115*	(0,018)	0,055*	(0,016)
Children under 16	-0,012	(0,016)	0,002	(0,016)
Small town	-0,004	(0,017)	0,010	(0,016)
Large town	-0,011	(0,017)	-0,021	(0,017)
Years of schooling	0,010*	(0,003)	0,012*	(0,003)
Full-time employed	-0,103*	(0,041)	-0,044	(0,035)
Part-time employed	-0,131*	(0,065)	-0,020	(0,037)
Unemployed	-0,067*	(0,029)	-0,033	(0,028)
Blue collar worker	0,107*	(0,041)	0,033	(0,036)
White collar worker	0,117*	(0,043)	0,025	(0,035)
Self-employed	-0,031	(0,046)	-0,065	(0,046)
Health job	-0,063	(0,080)	0,108*	(0,038)
Civil servant	-0,055	(0,055)	0,075	(0,061)
Net-household inc./1000	0,009	(0,005)	-0,004	(0,004)
West Germany	0,092	(0,057)	0,110*	(0,055)
Life Satisfaction	0,033*	(0,004)	0,027*	(0,004)
Oral Interview	0,132*	(0,013)	0,129*	(0,013)
Long Interview	-0,103*	(0,013)	-0,093*	(0,013)
Constant	-0,507*	(0,129)	-0,852*	(0,127)
Log-pseudolikelihood	-23727.056		-23994.092	
Observations	82,621		88,575	

Star indicates significance at the 5% level; Standard errors in parentheses

Table A2: Variable description and sample means

Variable	Mean Males	Mean Females	Description
SHI	0.88	0.93	Dummy for statutory health insurance
Health Satisfaction	6.78	6.58	Self-assessed health satisfaction betw. 0 (very bad) and 10 (very good)
Hospital Stays prev. Year	0.33	0.37	Number of hospital stays in the previous year
Disabled	0.13	0.09	Dummy for disability
Sport	2.16	2.00	Doing sports betw. 1 (almost never) and 4 (at least once a week)
Unemployment Rate	10.85	10.87	Average national unemployment rate per year
Absent > 3 days	0.32	0.25	Dummy for being absent for more than 3 days
Age	45.06	46.30	Age
Married	0.67	0.63	Dummy for being married
Children under 16	0.34	0.35	Number of children under 16
Small town	0.42	0.41	Dummy for living in a town with less than 20.000 inhabitants
Large town	0.32	0.33	Dummy for living in a town with more than 100.000 inhabitants
West Germany	0.77	0.77	Dummy for West Germany
Years of schooling	11.71	11.26	Years of Schooling
Full-time employed	0.63	0.28	Full-time employed
Part-time employed	0.01	0.15	Part-time employed
Unemployed	0.07	0.07	Unemployed
Blue collar worker	0.31	0.12	Blue collar worker
White collar worker	0.22	0.29	White collar worker
Self-employed	0.07	0.03	Self-employed
Health job	0.01	0.03	Health job
Civil Servant	0.05	0.02	Civil Servant
Net-household inc./1000	2.55	2.38	Net-household inc./1000
Life Satisfaction	7.00	7.00	Self-assessed life satisfaction between 0 (very bad) and 10 (very good)
Oral Interview	0.45	0.48	Oral Interview
Long Interview	0.52	0.50	Long Interview
Number of observations	82,621	88,408	