The effects of expanding the generosity of the statutory sickness insurance system ‡

Nicolas R. Ziebarth
SOEP at DIW Berlin and TU Berlin*

Martin Karlsson
TU Darmstadt**

October 1, 2010

Abstract

This article analyzes the effects of an increase in statutory sick pay from 80 to 100 percent of forgone gross wages in Germany. Difference-in-Difference analyses show that the increase in generosity decreased overall attendance in the private sector by 10 percent or 1 day per worker per year (20 percent or 2 days amongst those actually affected). Heterogeneity in response behavior was of great importance and employee health its main driver. For employers, the increased contribution represented increased labor costs of about €1.8 billion per year. Our empirical evidence supports the notion that employers tried to compensate for this shock to labor costs by increasing overtime and decreasing wages.

Keywords: sickness absence, statutory sick pay, generosity of social insurance, natural experiment, Socio-Economic Panel Study (SOEP)

JEL classification: H51; I18; J22; J32

‡We would like to thank Daniela Andrén, Tim Barmby, Mattias Bokenblom, Jörg Breitung, Stefano DellaVigna, Liran Einav, Christina Gathmann, David Granlund, Dan Hamermesh, Lars Hultkrantz, Guido Imbens, Per Johansson, Michael Lechner, Olivier Marie, Raymond Montizaan, Martin Olsson, Mårten Palme, Steve Pischke, Per Petterson-Lidbom, Peter Skogman Thoursie, Jan C. van Ours, Johan Vikström, and participants in seminars at the 2010 Annual Conference of the European Society for Population Economics (ESPE 2010), the 2010 Conference of the Applied Econometrics Association on Healthy Human Resources, the Workshop on Absenteeism and Social Insurance in Uppsala, the Joint Empirical Social Science (JESS) Seminar at the Institute for Social & Economic Research (ISER) at the University of Essex, and the Berlin Network of Labour Market Researchers (BeNA) for their helpful comments and discussions. The paper will also be presented at the 2011 meeting of the American Economic Association (AEA 2011) in Denver. We thank Felix Heinemann, TU Darmstadt, for excellent research assistance. All remaining errors or shortcomings of the article are our own. Nicolas R. Ziebarth gratefully acknowledges support from the *Stiftung der Deutschen Wirtschaft* (sdw, Foundation of German Business) in the form of a scholarship.

*Corresponding author: Nicolas R. Ziebarth, German Institute for Economic Research (DIW) Berlin, Socio-Economic Panel Study (SOEP), Mohrenstraße 58, 10117 Berlin, Germany, and Berlin University of Technology (TU Berlin), phone: +49-(0)30-89789-587, fax: +49-(0)30-89789-109, e-mail: nziebarth@diw.de

**Technische Universität Darmstadt, Marktplatz 15 - Residenzschloss, 64283 Darmstadt, e-mail: karlsson@vwl.tu-darmstadt.de

1 Introduction

Research in labor economics has long been preoccupied with how the social insurance system affects labor market performance; one need only think of the numerous studies on how unemployment insurance affects the behavior of the unemployed. In light of this, it seems odd that economists have devoted so little attention to a major form of social insurance that is directly linked to the labor market: sickness absence insurance. While statutory sick leave is almost unknown in the US and Canada, statutory sickness insurance is an integral part of social insurance systems in Europe.

Statutory sickness insurance protects employees against temporary income losses that arise from workplace absences due to illness. The United States has *Workers' Compensation Insurance (WCI)*, which covers incomes losses due to work-related sickness and is administered on the state level, and *Disability Insurance (DI)*, which replaces income losses stemming from long-term work absences due to disabilities and is administered on the federal level. What is relatively unknown, is that five states have forms of sickness insurance that are quite similar to those in Europe. These are referred to as "temporary disability insurance" or "cash sickness benefits." In 2005, the total sum of net benefits for temporary disability insurance in California amounted to \$4.2 billion, while the total sum for unemployment insurance amounted to \$4.6 billion (Social Security Administration, 2006, 2008).

Interestingly, a heated debate has emerged in the US over the last few years about the implementation of universal statutory sick leave on the federal level. A bill called the Healthy Families Act has been introduced in the House of Representatives as well as in the Senate. The bill foresees that every US employer with more than 15 employees would be required to provide sick pay for up to seven days per year. Many politicians as well as various lobbying groups strongly support the bill, arguing that it would increase employee productivity by reducing the rate of work attendance despite illness. The present paper contributes to this debate by illustrating potential effects of introducing or expanding a statutory sick leave scheme.

The literature on WCI and DI provides some empirical evidence. Two studies from the US have analyzed the impact of changes in benefit levels for WCI: Meyer et al. (1995) found that a 1987 increase in benefit levels led to longer duration of leave, while Curington (1994) presented mixed results based on data from the 1960s and 1970s. Issues surrounding DI have also attracted a great deal of attention in the recent literature. A number of studies have found that the generosity of DI affects labor supply decisions at the extensive margin (Bound, 1989; Gruber, 2000; Chen and van der Klaauw, 2008), although there is also evidence that this is not always the case (Campolieti, 2004). Researchers have also studied the DI application process (e.g., Burkhauser et al. (2004)) and the decision to apply for benefits within a lifecycle context (Chandra and Samwick, 2005). But compared to WCI and DI, sickness absence insurance systems cover a much broader range of illnesses and also provide benefits for short-term absences from work. Thus, although related, the empirical findings on DI and WCI are probably not directly applicable to sickness absence insurance.

Only a few studies have convincingly identified causal effects in the context of the sickness absence insurance. Using data from a large Italian company, Ichino and Maggi (2000) showed that cultural backgrounds determine temporary absence behavior to a large extent. In a recent paper, Ichino and Moretti (2009) showed that the menstrual cycle explains one-third of the workplace absence gap between men and women. While there have been many other studies on correlates of absence behavior, there is a paucity of empirical findings showing how the design of sickness insurance relates to absence behavior and the labor market. The literature contains only a handful of studies providing evidence on this relationship (Johansson and Palme, 1996, 2002, 2005; Henrekson and Persson, 2004; Puhani and Sonderhof, 2010), and almost all of them come from Sweden and are based on Swedish administrative data. Administrative data have various advantages over survey data but contain very little socio-economic information dealing, for instance, with individual health. Moreover, since these studies examine changes in sick pay levels that apply to every employee in Sweden without exception, they rely on before-after estimators, making it difficult to disentangle reform effects from general time trends. However, all of the studies cited above find that employees adapt their short-term sick leave behavior to economic incentives.

The present study provides clear-cut evidence on how a substantial increase in statutory sick leave benefits causally has affected sick leave behavior in Germany. On January 1, 1999, German statutory sick pay was increased from 80 to 100 percent of foregone gross wages, making the sickness insurance system substantially more generous. German employers are required to provide statutory sick pay for a period of six weeks per illness, starting on the first day of the illness, without any further benefit caps. To estimate the effects of the reform, we use representative SOEP survey data on Germany, Europe's most populous country. Our identification strategy relies on a well-defined control group and the use of parametric, non-parametric, as well as combined difference-in-differences approaches. Moreover, it allows us to distinguish between "intent-to-treat" effects, which are relevant for policy makers, and the response in labor supply to changes in the replacement rate as actually implemented. In addition, we not only show how expanding the generosity of a social insurance system affects labor supply decisions on the intensive margin, but also attempt to unravel the mechanisms underlying these decisions. To our knowledge, this is the first paper to attempt such a unified analysis. Furthermore, we provide empirical evidence on how employers might have reacted to this shock in labor costs in a highly regulated labor market.

Based on evidence presented in the first part of the paper that the increased statutory sick pay has decreased employee attendance on average by 10 percent or 1 additional day per employee and year (corresponding to 2 days or 20 percent amongst those actually affected), we proceed in the second part as follows: first, we demonstrate that heterogeneity in the reform effects plays a crucial role. Our results show that the reform effect is driven mainly by employees in bad health. This finding supports arguments citing the impact of decreased "presenteeism," a term used mainly in the social sciences and medicine. Presenteeism, the opposite of absenteeism, occurs when employees go to work despite being sick. A decrease in presenteeism as well as an increase in absenteeism or shirking may both be plausible explanations for the increase in workplace absence. Typically, economists refer to any behavioral change that is triggered by a change in insurance coverage as moral hazard. Thus, this paper strongly supports the notion that moral hazard plays a substantial role in the use of sickness absence insurance.

In the second step, we test whether expanding the generosity of the sickness insurance improved employee health on the whole. We do not find any empirical evidence of such an effect, nor do we find evidence that work satisfaction changed after the reform. These findings are consistent with the view that shirking was the main mechanism underlying the behavioral responses.

In the final part of the paper, we look at the employers' side of the coin. Recall that in Germany—as well as in most other European countries—employers are required by law to provide statutory (short-term) sick pay. This obligation was expanded by the reform. We calculate that, as a result of the reform, labor costs increased by about €1.8 billion per year. This figure represents an annual increase in employer-provided sick pay costs of about 8 percent and is very close to what the German Federal Statistical Office (2001) reports based on administrative data. Thus, one would expect that employers reacted to such an exogenous shock to labor costs. However, the German labor market is highly regulated, and dismissal protection legislation there is among the strictest worldwide. By evaluating the dynamics of overtime hours and wages relative to unaffected occupational groups, we suggest in the last part of the paper that employers may have tried to pass on the increased labor costs by increasing overtime hours and decreasing wages.

2 The German Sickness Insurance System and Policy Reform

2.1 The Sick Pay Scheme and Monitoring System

Before the implementation of the new law, every German private-sector employer was legally obligated to pay 80 percent of foregone wages for up to six weeks per sickness spell.[1] Obviously, self-employed people are not eligible for employer-provided sick pay. Public sector employees and apprentices are guaranteed 100 percent sick pay for up to six weeks per sickness spell. Henceforth, we use the term *short-term sick pay* as a synonym for employer-provided sick pay and *short-term sickness absence* as a synonym for absences of less than six weeks due to illness.

In the case of illness, employees are required to inform their employer immediately about both the illness and its expected duration. From the fourth day of a sickness spell on, a doctor's certificate is required and is usually issued for up to one week depending on the illness. If the illness lasts more than six continuous weeks, the doctor must issue a certificate of long-term illness. From the seventh week onwards, sick pay is disbursed by the sickness fund and is reduced to 70 percent of foregone gross wages for those who are insured under the Statutory Health Insurance (SHI).

Monitoring is carried out primarily by the *Medical Service of the SHI*. One of the main objectives of the Medical Service is to monitor sickness absence. German social legislation requires the SHI to contact the Medical Service and request a medical opinion to resolve any doubts regarding the validity of sick leave claims. Such doubts may arise if someone is absent for short periods with unusual frequency or is regularly sick on Mondays or Fridays. Similarly, if a doctor certifies sicknesses with unusual frequency, the SHI may call for an expert assessment of that doctor. The employer also has the right to request an expert assessment by the Medical Service, which is based on medical records, workplace information, and a statement that the patient is required to submit. If necessary, the Medical Service has the right to conduct a physical examination of the patients and to cut their benefits.[2] In 2007, about 2,000 full-time

[1] The entitlement is codified in the *Gesetz ber die Zahlung des Arbeitsentgelts an Feiertagen und im Krankheitsfall (Entgeltfortzahlungsgesetz)*, article 3, 4. Sick pay is calculated based on regular earnings and not overtime work.

[2] The text of the laws can be found in the Social Code Book V, article 275, article 276.

equivalent and independent doctors worked for the medical service and examined 1.7 million cases of absenteeism (Medizinischer Dienst der Krankenversicherung (MDK), 2008).

2.2 The Policy Reform

In the election campaign of 1998, the Social Democrats and the Greens promised to increase statutory sick pay from 80 to 100 percent of foregone gross wages should they form a new coalition government. The announcement was a reaction to a cut in sick pay under the previous center-right government under Chancellor Kohl in October 1996. At that time, beside a reduction of long-term sick pay, short-term sick pay was decreased from 100 to 80 percent of forgone gross wages. Ziebarth and Karlsson (2010) analyze the effect of the cut in short-term sick pay and find that it reduced the average number of absence days by about 12 percent for employees with up to 5.5 absence days.[3] The majority of Germans perceived the cut in sick pay as unfair and socially unjust and a number of strikes were organized opposing it. Immediately after the election of the new center-left government in September 1998, a law was passed that went into effect on January 1, 1999, increasing statutory short-term sick pay from 80 to 100 percent of foregone gross wages.[4]

Although statutory sick pay was increased by 20 percentage points in 1999, it did not change conditions for all private-sector employees, since employers can voluntarily provide sick pay over and above the minimum requirements. This, incidentally, is always a problem when analyzing the effects of changes in statutory minimum requirements, i.e, *all* studies in this strand of the literature face this issue. After the cut in sick pay in October 1996, and partly in response to union pressure, employers from various sectors had agreed in collective wage agreements to continue paying 100 percent of wages during sick leave. There are no official figures on how many employees benefited from this, but in 1998, union leaders proudly declared that 13 out of 27 million employees would receive 100 percent sick pay (Jahn, 1998).[5] In 1997, a poll among craftsmen's businesses showed that 51 percent were voluntarily providing 100 percent sick pay, probably due to the close relationship and mutual trust between employers and employees in

[3] For various reasons, it makes sense to analyze the effects of a decrease in coverage separately from the effects of an increase in coverage: first, the effects may be expected to differ. Second and more importantly, the sick pay reform was accompanied by various other reforms that act as confounding factors in the estimation: for example, a waiting period for new hires was introduced, the basis of calculation was changed, and long-term sick pay was also cut. Moreover, treatment and control groups differ among all these reforms, which requires different identification strategies. In addition, Ziebarth and Karlsson (2010) do not provide evidence on either the underlying operating mechanisms or employers' responses to the reform. The estimated change in sick leave behavior is, however, in line with the findings of the present study.

[4] Passed on December 19, 1998, this law is the *Gesetz zu Korrekturen in der Sozialversicherung und zur Sicherung der Arbeitnehmerrechte, BGBl.I 1998 Nr. 85 S.3843-3852.*

[5] Both figures include around 3.3 million public-sector employees (German Federal Statistical Office, 1999).

these small companies (Ridinger, 1997). Like all of the other studies that have evaluated the impact of changes in statutory benefit levels, we assess the overall impact of the law among private-sector employees, comparing them with completely unaffected occupational groups such as the self-employed and public-sector employees. However, since compliance with the reform was incomplete, we also estimate the actual labour supply response to the increase in the replacement rate in industries where the degree of compliance is known.

Even with 80 percent statutory sick pay, Germany provided among the most generous sick leave benefits worldwide. In 1998, the total sum of employer-provided sick pay amounted to €22.3 billion, exceeding 1 percent of GPD (German Federal Statistical Office, 2001). At that time, there were concerns among German economists that the high overall labor costs could be one of the main reasons for the persistently high unemployment rate in Germany. Germany was ranked among the top among OECD countries in total labor costs per hour. Since sick pay represents (non-wage) labor costs and functions like a tax on labor, the German Council of Economic Advisors disagreed with the increase of the minimum sick pay level and warned that it would pose a new obstacle to job creation (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung, 1998).

3 Data and Variable Definitions

For the empirical analysis, we use data from the German Socio-Economic Panel Study (SOEP). Aside from the SOEP, there is no other data set that includes representative information on sick leave in Germany. The SOEP is a household panel survey that began in 1984 and that focuses on labor market activities and earnings. It samples a rich array of subjective and objective workplace characteristics and socio-economic background information. Moreover, it includes self-reported attitudes of the respondents and personality traits. Further details can be found in Wagner et al. (2007).

For the main specifications, two pre-reform and two post-reform years are used; thus, we exploit information on sick leave for the years 1997 to 2000.[6] We restrict our working sample to respondents who are employed full-time and between 25 and 55 years of age. We do not use respondents with item non-response on relevant variables.

[6] Since current as well as retrospective information is sampled in every wave, we match the retrospective information with the current information for each year if the respondent was interviewed in both years. If not, we use the information available and assume that it has not changed from one year to the next.

3.1 Sick Leave Measure and Covariates

The SOEP offers detailed information about employment histories, job characteristics, type of job, and the various income sources. Information on self-assessed health, medical care usage, and the number of sick leave days is also sampled.

We call the dependent variable *Daysabs* and generate this count measure one-to-one from the answers to the following question: *"How many days off work did you have in 19XX [200X] due to illness? Please enter all days, not just those for which you had a doctor's certificate."* Relying on self-reported information rather than administrative data has both drawbacks and benefits. Clearly, the possibility of measurement errors is a significant drawback. The more periods of illness a respondent had in the previous year, the larger the recall bias is expected to be. In the case of underreporting, the estimated effects would be downward biased. Measurement errors inflate standard errors and lead to less precise estimates.

On the other hand, the overwhelming advantage of self-reported data over administrative data is that they provide a measure of the *total* number of days of sick leave. Researchers working with register data often face the problem that only doctor-certified sick leave is included, and that employer-provided sick leave is often left out. This almost always leaves the researcher with censored data and makes certain types of analyses impossible.[7] However, having an uncensored measure of the total number of days of sick leave based on survey data comes at the cost of not having detailed spell data.

The whole set of explanatory variables can be found in Table 1. The control variables used the main specifications (Part A of Table 1) are categorized as follows: the first group contains variables on personal characteristics such as the dummy variables *female, immigrant, East German, partner, married, children, disabled, health good, health bad, no sports*, and *Age* (Age^2). The second group consists of educational controls such as the degree obtained, the number of years with the company, and whether the person was trained for the job. The last group contains explanatory variables on job characteristics: among them are *blue-collar worker, white-collar worker*, the size of the company, and *gross wage per month*. Apart from including various interaction terms between these covariates and *years with company* as well as *gross wage per month*,

[7] Take the case of Sweden and the impact of changes in the waiting period: before 1987, Sweden had a waiting period with zero compensation for the first day of illness. In the 1990s, the waiting period and the employer-provided sick pay period were changed several times, generating a register base which is censored, and the censoring varies with the reforms (see Henrekson and Persson (2004) for more details). In addition to the absence of a natural control group, this makes it difficult to identify causal effects in the case of Sweden. Interestingly, as has been discussed in the introduction, almost all of the studies carried out so far on the relationship between sick pay levels and short-term sickness absence come from Sweden.

we also control for the annual state unemployment rate. In the parametric approaches, state dummies net out permanent differences across states and year dummies take account of common time shocks.

[Insert Table 1 about here]

For the extended analyses in the second part of the paper, we employ additional covariates (Part B of Table 1). These additional covariates either incorporate a substantial degree of item-non-response or were only collected in specific years, which is why we do not use them in the main specifications. Further details about these variables can also be found in the notes to Table 7.

3.2 Treatment and Control Group

In our baseline specification, the treatment group consists of all private-sector employees except apprentices. The control group incorporates public-sector employees, apprentices, and the self-employed—all those who did not experience a change in their sick pay levels during the period under consideration. The dummy *Treatment Group* has a one for those belonging to the treatment group and a zero for those belonging to the control group. In total, we have 15,140 observations in the treatment group and 7,918 observations in the control group.

In a second step, we then try to take imperfect compliance into account. In this part, we focus on industries for which the degree of compliance is known (cf. Ziebarth and Karlsson (2010) for an overview). The treatment variable is in this case a fraction, which represents the proportion of workers in an industry that get 100 percent sick pay.

4 Estimation Strategy and Identification

4.1 Assessing the Causal Reform Effects on Sick Leave

Parametric Approaches

OLS

We start by estimating conventional parametric difference-in-differences (DiD) models. Consider the following equation:

$$y_{it} = \lambda p99_t + \pi D_{it} + \theta DiD_{it} + s'_{it}\psi + \rho_t + \phi_s + \epsilon_{it} \tag{1}$$

where y_{it} stands for the annual number of days of sick leave for individual i in year t, $p99_t$ is a post-reform dummy, D_{it} is the treatment group dummy, and DiD_{it} is the regressor of interest. It has a one for respondents in the treatment group in post-reform years and gives us the causal reform effect should certain assumptions hold true. It can also be interpreted as the interaction term between the treatment group dummy and the post-reform dummy. By including additional time dummies ρ_t we control for common time shocks that might affect sick leave. State dummies ϕ_s account for permanent differences across the 16 German states along with the annual state unemployment rate that controls for changes in the tightness of the regional labor market and that is included in the $K \times 1$ column vector s'_{it}. The other $K-1$ regressors are made up of personal controls including health status, educational controls, and job-related controls as explained in Section 3.1. In addition to the covariates that are displayed in Part A of Table 1, we also include various interaction terms between them. As usual, ϵ_{it} stands for unobserved heterogeneity and is assumed to be normally distributed with zero mean. To begin with, equation (1) is estimated by OLS.

Zero-Inflated NegBin-2 (ZINB-2)

The number of days of sick leave is a highly skewed count variable with excess zero observations (about 50 percent of the sample) and overdispersion, i.e., the conditional variance exceeding the conditional mean. Hence, it is appropriate to fit count data models, which might capture the skewed distribution better than simple OLS regressions. Based on the Akaike (AIC) and Bayesian (BIC) information criteria and various Vuong tests, we found the so-called *Zero-Inflated Negative Binominal Model (NegBin)* to be appropriate for our purposes (Cameron and Trivedi, 2005). Thus, we use this model in all count data specifications.

Non-Parametric Approaches

A fundamental alternative to estimating parametric models is matching. In principle, matching intends to make treatment and control observations more comparable by assigning each treated unit one or more control units that are similar in terms of observable characteristics. Under the "conditional independence" or "unconfoundedness" assumption, which claims that, after having conditioned on observables, the treatment is independent of the outcome, the assignment to treatment can be interpreted as random—as if it were generated by a randomized experiment (LaLonde, 1986). Various matching methods exist.

Most matching analyses use the method of propensity score (PS) matching. Rosenbaum and Rubin (1983) have shown that conditioning on the propensity score (PS)—the probability of being selected into the treatment group—is equivalent to selecting pairs of treated and control observations based on every covariate dimension, provided that unconfoundedness holds. We estimate the PS by means of a logit model and select the covariates to be included out of the total number of covariates (Part A of Table 1) using likelihood ratio tests on zero coefficients. In a first step, we do this for control variables in levels and in a second step for their interactions (Imbens, 2008).

In addition to a plausible selection on observables story, matching requires that the distributions of the covariates for treated and control observations overlap to a large extent. In this setting, the common support assumption is fulfilled, as can be seen in Figure 1. The PS distribution for both groups shows a large overlap with the region of common support lying between PS values of 0.05 and 0.92.

[Insert Figure 1 about here]

The first non-parametric method that we employ is stratification matching or blocking. Based on the difference-in-differences indicator, DiD, the sample is cut into blocks such that the propensity score is balanced within each block. Then, block-by-block average treatment effects on the treated are obtained by taking the difference between the average outcome for $DiD = 1$ and $DiD = 0$ within each block. Afterwards, the overall treatment effect on the treated can be computed as the weighted average of the block-by-block treatment effects (Rosenbaum and Rubin, 1984). Cochran (1968) has shown that, in linear models, five blocks are sufficient to reduce the bias that is associated with the overall simple outcome difference between treated and untreated samples by more than 95 percent.

The second method is k-to-one nearest neighbors matching with replacement. Again, based on the difference-in-differences indicator, DiD, the propensity score is estimated and to every observation with $DiD = 1$, the k most similar observations with $DiD = 0$ are assigned. Then, the outcome difference of each pair is taken to compute the average treatment effect on the treated (Heckman et al., 1998; Lechner, 2002).

Combining Parametric and Non-Parametric Approaches

Both regression and matching methods have drawbacks. If treated and control units differ substantially in their observed characteristics, then parametric approaches use the covariate distribution of the controls to make out-of-sample predictions.

Imbens and Rubin (2009) propose to evaluate differences in covariates for treatment and control group by the scale-free normalized difference:

$$\Delta s = \frac{\bar{s}_1 - \bar{s}_0}{\sqrt{\sigma_1^2 + \sigma_0^2}} \tag{2}$$

with $\bar{s}_1$ and $\bar{s}_o$ denoting average covariate values for the treatment and control group, respectively. σ stands for the variance. As a rule of thumb, a normalized difference exceeding 0.25 is likely to lead to sensitive results (Imbens and Wooldridge, 2009).

Applied to our case, we first look at how the covariate distribution for the treatment group differs in comparison to the control group, i.e., we compare private-sector employees to those whose sick pay was not affected throughout the whole period under consideration. Table 2 shows in column (1) the means of the covariates for the treatment group and in column (2) the means of the covariates for the control group. It appears that the two groups are very similar with respect to their observable characteristics. This presumption is reinforced by column (3), which displays the normalized difference. Indeed, all of the values are smaller than 0.20 and some tend towards zero.

[**Insert Table 2 about here**]

We now apply two different matching procedures to improve the balancing properties across treatment and control group. Using combined matching and regression approaches requires this as a first step. In the second step, we then apply regression approaches to these matched samples. Note that the first step—balancing covariate distributions—requires that we match on the treatment group indicator D, not on DiD.

Columns (3) to (6) show the "matched sample" and the covariates' mean values for the treatment and control group plus the normalized difference. We obtain this sample using five-to-one nearest neighbors matching. The matched sample shows better balancing properties than the raw sample and all normalized differences are below 0.15.

Using stratification matching, we obtain the "blocked sample" in columns (7) to (9). It is easy to see that blocking substantially improves the balance of the covariates between the treatment and the control group. The normalized difference for almost all covariates yields values of less than 0.1.

However, even for the matched and the blocked samples, small differences between treatment and control group remain. These differences may lead to biased estimators. Abadie and Imbens (2007) have shown that the simple nearest neighbor matching estimator includes a bias term, which leads to inconsistencies and should be corrected for.

Imbens and Wooldridge (2009) propose two approaches that both combine the strengths of parametric and non-parametric estimators; both estimators work well in practice, and both estimators lead to robust results. Approach number one is a combined blocking and regression approach. In the first step, stratification matching is applied to the raw sample to obtain a blocked sample with better balancing properties, as in columns (7) to (9). In the second step, parametric regressions—as detailed in Section 4.1—are run within each block. Then, the within-block treatment effects are weighted by the relative size of the blocks and aggregated into an overall average treatment effect on the treated. The crucial point is that the covariate distributions within each stratum are very similar and, thus, out-of-sample predictions are avoided.

The second approach also aims to smooth differences in covariates between treatment and control group and additionally corrects for the bias described in Abadie and Imbens (2007). It combines regression and k-nearest neighbors matching. In the first step, using only the untreated who were matched to the treated, we conduct a linear regression of the outcome on the covariates.[8] Then, in the second step, the counterfactual potential outcome for the case without treatment, y^0, is calculated as (Abadie et al., 2004):

$$\hat{y}_i^0 = \begin{cases} y_i \text{ if DiD} = 0 \\ \frac{1}{M}\sum_{j\epsilon\Gamma_M}(y_j + \hat{\psi}_0'(s_i - s_j)) \text{ if DiD} = 1 \end{cases}$$

where Γ_M denotes the set of indices for the M closest matches for unit i, and y_j is the outcome which is matched to unit i.

[8] Here, a linear model is used. However, various specifications are conceivable.

Identification of Causal Effects

In the previous subsections, we discussed how we utilized the rich set of socioeconomic background information to make the treatment and the control group as comparable as possible. As can be seen in columns (6) and (9) of Table 2, both matching and blocking yield two samples that are almost identical in terms of observables.

However, the crucial identifying assumption in any difference-in-differences (DiD) analysis is that all relative post-reform changes in the outcome variable of the treatment group can be traced back to the reform. In other words, it is assumed that—conditional on all personal, educational, and job characteristics, as well as time and year dummies—there are no unobservables that impact the dynamic of the outcome differently for the two groups. This common time trend assumption is not directly testable. However, we believe that it is very likely to hold in our context.

First, we are analyzing a reform that applied to a large and well-defined group in the labor market—private-sector employees. The reform was implemented at the federal level and reduced the cost of workplace absence, an outcome that we are able to observe directly. Since the reform was an automatic reaction to a previous reform, it was also exogenous in the sense that it was not implemented to combat rising absenteeism but to keep a promise made during an election campaign (Besley and Case, 2000).

Second, we can exclude the possibility that selection into or out of the treatment contaminated our estimates since we rely on panel data and can identify job changers. For example, we can test the robustness of our results with respect to sample composition changes over time and labor market attrition. As a robustness check, we restrict the sample to respondents, whom we observed as working in the pre- *and* post-treatment period and who answered the questionnaire without item non-response. In addition, it might be that, in response to the reform, public-sector employees and the self-employed applied for jobs in the private sector, where working conditions had improved. It might also be that the increase in sick pay induced more non-working people to accept jobs in the private sector. In the robustness checks, we can tackle such selection concerns by excluding people who changed jobs or sectors.

Third, we are able to control for a rich set of background variables like job characteristics and (self-reported) health status. The latter is by far the most important determinant of sickness absence.

Fourth, since it is possible to indirectly test the plausibility of the common time trend assumption, we present the results of placebo regressions. Placebo regressions assume that the reform analyzed took place in a year without any other reform. Should the coefficient of interest be significant in a non-reform year, the common time trend assumption would be seriously challenged.

Finally, in Figure 2, we display the average number of sickness absence days for several pre- and post-reform years and both groups. In 1996, as explained in Section 2, various sick leave reforms were implemented that all affected subsamples that differ from those analyzed here: thus, we can only use the pre-reform years of 1997 and 1998 for this exercise. However, we plot the absence rates for five post-reform years, which should also yield enough evidence of the plausibility of the common time trend assumption. Since no other sick leave legislation was passed after 1999, a priori, we would expect to see a jump in the number of days of sick leave for the treatment group in the reform year 1999, but more or less parallel time trends in subsequent years. This is exactly what we find. We observe relatively parallel curves for both groups in the pre-reform years. After the reform went into effect in 1999, the absence curve for the treatment group shifts upwards and subsequently runs parallel to the curve for the control group. In this graph, it seems as if the reform effect lasts for about four years, since we then observe a closing of the gap in absence days. However, it should be kept in mind that Figure 2 paints a raw, unconditional picture. One explanation for the closing of the gap from 2001 to 2002 could be that unemployment rates increased by five percent or 0.5 percentage points (200,000 unemployed) between the two years—after they had been decreasing more or less monotonically for four years, i.e., since 1997 (German Federal Statistical Office, 2009). It is a well-documented stylized fact that changes in unemployment and absence rates are negatively correlated (Askildsen et al., 2005). While the figure here represents only descriptive evidence, we also correct the sample composition with respect to a rich set of covariates in the empirical assessment below.

[**Insert Figure 2 about here**]

In recent years, the drawbacks and limitations of DiD estimation have been debated extensively. A particular concern is the underestimation of OLS standard errors due to serial correlation in the case of long time horizons as well as unobserved (treatment and control) group effects (Bertrand et al., 2004; Donald and Lang, 2007; Angrist and Pischke, 2009). To cope with the serial correlation issue, we focus on short time horizons. In addition, to provide evidence on whether unobserved common group errors might be a serious threat to our estimates, in robustness checks, we cluster on the state×year ($16{\times}4 = 64$ clusters) as well as on the industry×year ($= 242$ clusters) level, where negotiations about the application of the reform took place (Angrist and Pischke, 2009).[9]

As has been discussed in Section 2.2, even before the increase in statutory sick pay, some employers agreed in collective bargaining to provide 100 percent sick pay voluntarily. We cannot precisely identify employees who were subject to such collective wage agreements. Our approach is to combine two different strategies. First, we focus on intention-to-treat effects, which is always required when analyzing changes in statutory minimum standards: in this strand of the literature, the *overall* effects of changes in statutory sick pay are evaluated. In contrast to other countries like Sweden, where differences in the labor agreements are more fragmented, polls for Germany at the time of the reform suggest that around half of all private-sector employees received statutory (80 percent) sick pay and the other half received 100 percent sick pay (Ridinger, 1997; Jahn, 1998). In an alternative specification, we focus on industries for which we know the degree of implementation of the reform. We still cannot identify exposure to treatment at the individual level, but we know the fraction of affected individuals in some industries. Since we use a linear model, this alternative approach—using the fraction of treated in each industry—provides us with an estimate of the effects of the expansion at the individual level.

4.2 Assessing Heterogeneity and Further Reform Effects

The previous subsections discussed the methods and assumptions for identifying how the generosity expansion causally affected sick leave behavior. In the second part of the paper, we take a step further and try to unravel some of the mechanisms at work behind the pure labor supply effects on the intensive margin. To our knowledge, all existing studies stopped at analyzing reform effects on workplace absence. We believe that much more can be learned in such a setting, especially when we take this natural experiment as an example of how social insurance and the labor market interact. It may be worthwhile to study interrelations between sickness absence insurance and the labor market that were triggered by the exogenous variation in the costs of absence to both employers and employees. In addition, effect heterogeneity is likely to be of high relevance in this setting.

We start by examining whether heterogeneity in the response to the policy change plays a role. A priori, one would expect that the reform effect is not uniformly distributed across socio-economic and workplace characteristics. For example, by differentiating the reform effects

[9] In the analysis of absence behavior by degree of implementation, we go one step further and cluster standard errors at the highest possible level.

by the health status of the respondents, we provide evidence on whether changes in employee sick leave behavior is attributable primarily to shirking or to presenteeism. Technically, we assess treatment effect heterogeneity by interacting possible covariates with the regressor *DiD* in equation (1). Then, we add this additional interaction term to the model.

In the next step, we attempt to understand how the reform has affected employers—whether directly, through higher non-wage labor costs due to increased sick pay levels, or indirectly, through an increase in workplace absence. Since we have individual-level information on days of sick leave and gross wages, and since we make use of SOEP frequency weights, we are able to calculate how much labor costs have increased. We can then attempt to empirically assess how employers might have reacted to this increase in labor costs, i.e., whether working hours, workplace climate, or wages in the private sector have changed relative to the unaffected occupational groups.

5 Empirical Results

A detailed discussion on the implementation of the various empirical approaches, their underlying assumptions, and the identification strategy can be found in the previous section. This section presents and discusses the main empirical results. In the next section, we first show how increasing the generosity of the sickness insurance system has causally affected sick leave behavior, and then provide evidence on the underlying mechanisms and further spillover effects.

5.1 Assessing the Causal Reform Effects on Sickness Absence

Parametric Approaches

We start by estimating parametric OLS and ZINB-2 models on the raw sample with all covariates of Part A of Table 1 included. In the following, we always display marginal effects. The parametric DiD estimates are displayed in columns (1) and (2) of Table 3. The OLS model yields an estimate of 1.366 that is statistically significant at the 5.4 percent level. The ZINB-2 model gives an estimate of 1.018 with a standard error of 0.468. The unconditional double difference of the means of the two groups for the two time periods is 1.441 (std. err. 0.732; not shown) and very close to the OLS estimate in column (1), which reinforces the credibility of the common time trend assumption.

[Insert Table 3 about here]

Non-Parametric Approaches

Columns (3) and (4) give the results when two different matching variants are applied using the matched subsample of Table 2. In the first step, the matched subsample of Table 2 is obtained by five-to-one nearest neighbors matching on the probability of belonging to the treatment group vs. the control group. As can be seen in column (6) of Table 2, the covariate distribution is almost perfectly balanced between the treatment and control group in this sample. In the second step, we perform two matching methods as described in the methods section. In this respect we could also call our non-parametric applications "two-step matching procedures." While we conduct matching on the probability of belonging to the treatment group to obtain the matched sample, afterwards, we perform matching on the difference-in-differences indicator. In other words, in the first step, we match on the treatment group dummy D, and in the second step, we match on the difference-in-differences indicator, *DiD*.

Hence, both columns (3) and (4) make use of the matched sample as shown in Table 2. In column (3), the underlying matched sample is stratified into blocks based on the propensity score (PS) that *DiD*=1. Then, by taking the average values of treated and untreated within each block, the block-specific reform effects are calculated, which are finally aggregated to a weighted overall average. This method gives an estimate of 1.138 with a standard error of 0.406. Column (4) yields the estimate when five-to-one nearest neighbors matching is applied to the difference-in-differences indicator using the matched sample. The estimated reform effect is 1.120 and significant at the one percent level.

Combining Parametric and Non-Parametric Approaches

According to Imbens and Wooldridge (2009), the most suitable methods combine regression and matching and are thus more flexible and robust than other methods. Column (5) shows the result when the raw sample is first stratified on the probability of belonging to the treatment group (hence it makes use of the blocked sample in Table 2) and then regressions as in equation (1) are run block-by-block. The overall treatment effect, which is 0.8973 and significantly different from zero, is obtained as an average of the within-block estimates weighted by the block size. The method used in the last column also combines matching and regression and eliminates a bias that has been proven to exist for nearest neighbor matching. More details can be found in Section 4.1 and Abadie and Imbens (2007). The resulting estimate is similar to the one in column (5) and yields a reform effect of 0.9986 (std. err. 0.33307).

It is remarkable that all estimates differ only slightly in size and that all point estimates are statistically different from zero and carry the expected sign. The size of the coefficients varies between 0.90 and 1.37 and the confidence intervals largely overlap. These findings suggest that the identified effect is very robust and not very sensitive to the functional form imposed. All in all, the conventional and transparent OLS difference-in-differences model does a relatively good job of estimating the effect of the reform. Thus, in the following, we will focus on conventional OLS models.

If we take the mean number of absence days in the pre-reform period for the treatment group, which was 9.7, and relate an estimate of 0.97 to it, we would conclude that the increase in statutory sick pay led to a 10 percent increase in the average number of absence days among the treatment group. As has been discussed previously, among the reform's target group, about half of all employees effectively experienced an increase in sick pay. Hence our estimates suggest that the employees actually affected increased their days of sick leave by about two days per year. Indeed, this assumption is almost exactly verified in the "full compliance estimation approaches" presented in the sext subsection.

Alternative Approaches Using Industry-Level Compliance

We now proceed to test specifically by how much the increase in sick pay increased the average number of sick days among those who actually experienced an increase. For that purpose, we reviewed all collective agreements in the main industries and identified industries that fully complied to the statutory minimum standards and provided 80 percent sick pay in the pre-reform years as well as those where only workers not covered by collective agreements were affected. Likewise, we were able to identify industries that provided 100 percent sick pay througout the entire period under consideration. The analyses in Table 4 are based on the fraction of full compliers within each industry, which we use as treatment variable. More details are provided in the notes to Table 4.

[Insert Table 4 about here]

The last three columns of Table 4 are entirely based on private sector employees, whereas the other unaffected occupational groups are included in the control group in the first three columns. Columns (1) and (4) make use of the standard "intent-to-treat" approach as above but are based on the same samples as the models in the other columns. The results for these columns are displayed for comparative purposes. Since not every employee in the treatment group in columns (1) and (4) was affected by the reform, we expect the estimated behavioral effects to be smaller than the ones in the other columns. The results from columns (2) and (3) as well as from columns (5) and (6) show clearly that the estimated "full compliance effect" of the expansion is around 2 days (corresponding to an increase in absence days of about 20 percent) and thus larger than the effect of the "intent-to-treat" approach. The implied arc elasticity with respect to the increase in the replacement rate would be 0.9. This finding is comparable with the results of the few existing studies that analyze similar reforms (Johansson and Palme, 2005).

Given the political economy of the reform and the notion that about half of all respondents in *Treatment Group* were effectively affected by the increase in sick pay, these estimates fit nicely to the results in Table 3. In addition, considering that Table 4 provides us with two new distinct approaches comparing alternative treatment and control groups, the results are remarkably robust.

Robustness Checks

For the sake of simplicity and to guarantee a sufficiently large sample size, henceforth, we focus on OLS DiD-models and our basic "intent-to-treat" approach. Apart from having analyzed the sensitivity of the results with respect to various parametric, non-parametric as well as combined methods, further robustness checks are shown in Table 5. Column (1) displays the result for a model that includes the lagged level of the total number of absence days as an additional covariate. This specification yields a positive and highly significant reform estimate of 1.568.

Column (2) checks whether panel or labor market attrition might drive our results. Only those who were observed working in the pre- *and* post-reform period at least once are included in the sample. By restricting the sample as such, we lose approximately 8,000 observations, and the precision of the estimate decreases; the estimate is only significant at the 11.8 percent level. However, we find a reform effect that is of similar size to the one in the main specification in column (1) of Table 3. When applying the Abadie and Imbens (2007) nearest-neighbors matching regression method to that specification, precision increases substantially, and the effect is significant at the one percent level (result not shown).

Columns (3) and (4) deal with concerns that selection into occupations might produce or bias the results. Column (3) excludes respondents who answered the following question with "yes:" *"Did you change your job or start a new one after December 31, 199X?"* We thereby capture job changers who might have selected themselves into (or out of) the treatment. The size of the estimate is almost identical to the main estimate in the first column of Table 3 and is marginally significant at the ten percent level. This also holds for column (4), where we provide an alternative robustness check on selection effects. In column (4), all private-sector employees who changed industry sector in the post-reform period are excluded from the sample. Given the reform design, it is likely that collective bargaining assured that sick leave regulations only varied across but not within industries. In another check, we look at whether the reform was followed by a change in the rate of job change. There is no evidence that this occurred. Between January 1997 and the date of the 1998 SOEP interview (most are conducted during the first three months of the year), 16.45 percent of all interviewees had changed jobs. This rate remained almost exactly the same for the period between January 1999 and the 2000 SOEP interview, namely 15.78 percent. In addition, looking at whether the distribution of job-changers across health states changed after the reform provides no such evidence either. From 1997 until the 1998 interview, 14.43 percent of all employees in poor or bad health changed jobs. From 1999 to 2000, the rate was 13.72 percent. As a final check, we look at whether the rate of changing occupations—i.e., between private sector, public sector, and self-employment—changed after the reform. From 1997 to 1998, 1.76 percent of all employees switched from the public to the private sector, up from 1.70 percent between 1999 and 2000. During the same two time periods, 0.44 and 0.37 percent, respectively, switched from self-employment to the private sector.

[Insert Table 5 about here]

Columns (5) and (6) of Table 5 cluster standard errors at the state×year (64 clusters) as well as at the industry×year (242 clusters) level to provide evidence on whether the group structure might be a serious issue in this setting. We find no evidence that this is the case. The plain standard error for the main model is 0.6758 (not shown). Clustering on the individual level slightly increases the standard error to 0.7097 (Column (1) of Table 3). Clustering on the state×year level yields a standard error of 0.5992 and clustering on the industry×year level yields a standard error of 0.6857.

The last column of Table 5 tests whether there is evidence that the increase in statutory short-term sick pay had any effect on the incidence of long-term absenteeism. The estimated coefficient is almost zero in magnitude and not significant; thus it is reasonable to conclude that the distribution of long-term absence spells remained stable after the reform.

As has already been mentioned, an indirect method to test the common time trend assumption is to perform the same analyses for years with no reform. Significant reform estimates for years with no reform would cast doubts on the assumption of no unobserved year-group effects. In this context, however, this is not the case as Table 6 demonstrates.

[Insert Table 6 about here]

5.2 Assessing Effect Heterogeneity and Further Reform Effects

Treatment Effect Heterogeneity and Health Effects

Table 7 displays extensive tests on treatment effect heterogeneity. Every column shows one OLS difference-in-differences model as in the main specification in column (1) of Table 3. The only difference is that the corresponding variable—with which we want to perform the heterogeneity test—is included both in levels and in interaction with the *DiD* regressor. Take as an example the first column of Panel A in the table. Here we want to check whether men have reacted differently from women to the reform. Hence, in addition to the gender dummy that was included in the model anyway, we interact the dummy variable *female* with *DiD* and run the model under the inclusion of this additional interaction term. The usual *DiD* point estimate tells us how men reacted to the reform. We find a highly significant 1.65 *DiD* estimate, suggesting that males reacted disproportionately to the reform. This is reinforced by the *DiD*×*female* point estimate. While the coefficient is imprecisely estimated, it is negative and large in magnitude. This provides some evidence that women did not react as strongly to the increase in sick leave benefits as men did, although the difference between men and women is imprecisely estimated.

Panel A tests heterogeneity in the response behavior to the reform with respect to six variables that we subsume under the category of "personal characteristics." We have already discussed the findings for gender. Interestingly, there is no evidence that the age or education matters in terms of how employees reacted. There is some evidence that the richer half of the population reacted less than the poorer half, although the difference is not statistically significant. In contrast, there is strong and statistically significant evidence that the bulk of the behavioral effect is driven by employees with a spouse or partner. One explanation could be that the utility from spare time is higher for employees with a partner.

Panel B exploits six (self-reported) health measures: self-assessed health (SAH), health satisfaction, a question on whether respondents feel impaired in their everyday tasks by their health status, and certified disability. Precisely how the six dummy variables are generated is explained in the notes to Table 7. The empirical results show that employees in bad health reacted much more strongly to the reform than the rest. Workers with disabilities, those with low health satisfaction, and those with low self-assessed health were induced by the reform to use between four and nine additional days of sick leave. This is a huge effect as compared to one day for the average population. By contrast, we find some evidence that healthy employees reacted less to the increase in insurance coverage than the average employee. There is even evidence that employees with very high health satisfaction did not react at all. This would be strong evidence against shirking—at least for this specific subgroup. Although the *DiD×high health satisfaction* coefficient in column (3) is only marginally significant, all three models that test the effects for healthy employees have negative signs on their interaction terms. Restricting the sample to respondents who indicated the best SAH category (equivalent to column (1)) and running the standard OLS-DiD model gives us an imprecisely estimated reform effect of 0.8. Using only respondents who were highly satisfied with their health yields an insignificant reform effect of 0.06 (equivalent to column (3)). Those who did not feel impaired by their health have an OLS-DiD coefficient of 1.01, which is significant at the five percent level, but still substantially lower than the one in column (1) of Table 3.

[**Insert Table 7 about here**]

In Panel C we exploit subjective workplace characteristics. In a statistical sense, we do not find evidence that the reform effect differs when the effect is stratified across these variables. However, the signs and sizes of the triple interaction terms are within what one would expect. There is some evidence that those with low job satisfaction took more days off, whereas those who were very worried about their job security or who were likely to lose their jobs within the next two years took fewer days off than the rest.

Panel D makes use of the rich panel data in another way, by looking at attitudes and personality traits of the respondents. Although insignificant, the triple interaction coefficient for respondents who felt that sickness should be insured privately is positive and of high magnitude, which is surprising. One might also find it surprising that those who claimed that "one needs to work hard for success" seem to have taken more days off than those who did not agree with this statement. Likewise, those who held the view that they can influence and have control over their life (columns (4) and (5)) seem to have reacted more strongly to the increase in sick pay.

All in all, we find strong evidence of a substantial degree of heterogeneity in responses to the increased generosity of sickness insurance. Although many effects are imprecisely estimated, the signs and sizes of almost all coefficients are close to what one would intuitively expect. A key finding is that employees in bad health reacted much more strongly than the population average. In contrast, healthy employees reacted at a below-average rate, and there is even evidence that some might not have reacted at all. In any case, the health status was the key driver of the change in sick leave behavior with respect to the decrease in absence costs. We find mixed evidence for the notion that shirking was primarily responsible for the decrease in employee attendance. On the one hand, some models show that healthy employees also changed their sick leave behavior—although not by as much as the average employee. Moreover, there is some support for the argument that employees who were dissatisfied with their jobs responded more strongly than the rest. On the other hand, we find that primarily unhealthy employees changed their behavior, and that those who were very satisfied with their health did not change their behavior at all.

If that presenteeism was widespread prior to the reform, it is possible that increasing insurance coverage decreased the fraction of employees who went to work despite being seriously ill. The finding that primarily employees in poor health were responsible for the increase in workplace absence is very much in line with this explanation. However, if this were indeed true, then one might also expect to find an improvement in employee health. We provide evidence on this by running the same OLS-DiD models as before, but using different measures of poor health as the outcome variable. The results are shown in the first three columns of Table 7. There is absolutely no evidence that the health status of employees has improved as a result of the expansion of the public insurance coverage. All estimates are very close to zero and insignificant.

[**Insert Table 8 about here**]

Moreover, as can be seen in column (4), we do not find that job satisfaction, as reported by employees, has changed in response to the reform.

Reform Induced Increase in Labor Costs

While until now, we have provided a great deal of empirical evidence and discussion on what might have happened on the employee side, we have completely ignored the employer side of the coin. Now we want to present empirical evidence on how expanding a social insurance system might affect firms and induce changes in the organization of and demand for work.

First, we assess how the increased obligation to provide sick leave benefits might have affected labor costs directly and indirectly. For the moment, we assume the world to be static. Then, the maximum overall increase in labor costs can easily be calculated by comparing the total employer-provided sick pay in the pre-reform years 1997/1998 with the total benefits in the post-reform years 1999/2000 under the assumption that every employer only provided the statutory 80 percent sick pay in the pre-reform years.[10] Thus, we calculate annual sick leave benefits for every employee in the sample and apply frequency weights to the sum. For the pre-reform period, we assume a replacement level of 80 percent of foregone gross wages and for the post-reform periods, we assume a replacement level of 100 percent. The frequency-weighted benefit sums for both periods are multiplied by the frequency-weighted number of employees in the treatment group. By taking the difference between pre- and post-reform years, we obtain a total maximum increase in labor costs of €5.153 billion for the two post-reform years.

This total increase in labor costs can be decomposed into three components. The first component is the intramarginal effect associated with the increase of the statutory sick pay level for the first six weeks from 80 to 100 percent of foregone gross wages. We approximate this amount by comparing the total sick leave payments in the pre-reform period to hypothetical sick leave payments for the same period and the same individuals, assuming that the sick pay was already increased to 100 percent at that time. We thus disentangle the direct labor cost effect from the effect that is induced by increasing absence rates as a consequence of the reform. Again, we do not need any regression results for this exercise and use the full sample. Our calculation yields a direct labor cost effect of €3.87 billion for both years. If we assume that half of all firms had already provided 100 percent sick pay before the reform, this direct effect reduces to €1.93 billion.[11]

The second component represents the indirect labor cost effect, which was triggered by the reform-induced increase in workplace absence. From Table 3, we infer that the overall reform-induced increase in absence days equals approximately one day. Hence, we take the average daily gross wage in the pre-reform years and multiply it by the frequency-weighted number of employees in these years, resulting in an indirect labor cost effect of €1.61 billion. If we assume

[10] For this overall calculation, we do not need any of the regression results. This is a simple descriptive exercise, in which we make use of the full sample, i.e., we consider all employees in the private sector who are between 18 and 65 years old. For employees who claimed that they had a long-term absence spell of more than six weeks, we set the value for total absence days to 42, as only the first six weeks of sick leave are paid by the employer.

[11] We, thereby, implicitly assume that employees who worked in firms that voluntarily provided 100 percent sick pay did not differ systematically in terms of absence days and wages from those who worked in firms that only provided statutory sick pay. This assumption is unlikely to hold. Thus, we probably overestimate the increase in labor costs.

that the increase was 0.9 or 1.1 days, we get indirect effects of €1.45 and 1.77 billion over the two years, respectively.[12] The residual is the third component which is caused by time trends, changes in wages, and changes in the employment structure.

The total reform-induced increase in labor costs is thus $(1.93 + 1.61)/2 = €1.77$ billion per year.[13]

We cross-check the plausibility of our labor cost calculations by looking at administrative data. The German Federal Statistical Office (2001) provides administrative data on the total sum of employer-provided sick pay for the whole of Germany, including voluntary sick pay and time trends. Our calculations are very much in line with the official data. According to the German Federal Statistical Office (2001), the total sick pay sum in 1998 was €22.9 billion and increased by €1.87 billion to €24.78 billion in 1999.[14] Note that our estimate of €1.77 billion is net of time trends and assumes that 50 percent of all private-sector employees were actually treated. On the one hand, the similarity of our figure to that from the Federal Statistical Office suggests that the SOEP is very accurate in sampling wages and absence information. On the other hand, it also provides indirect evidence of the plausibility of our identification strategy and the assumption that about 50 percent of all private-sector employees were affected by the reform.

Relating our calculated—reform-induced—increase in annual labor costs to the total employer-provided sick leave benefit sum for 1998 yields an increase in sick leave costs of 7.7 percent. Using official numbers, including time trends, we end up with an increase of 8.2 percent.

Empirical Evidence on Employers' Attempts to Compensate for Increased Labor Costs

Since employers maximize profits, they must have responded in some way to the exogenous increase in labor costs of about €1.8 billion per year. In Germany at that time, very high total labor costs—especially in an international comparison—were a matter of serious concern for politicians, economists, and employers. These high labor costs were claimed to be the main barrier to job creation in Germany. Various researchers studied the relationship between labor

[12] Here, we focus on the same data set that we use to obtain the estimated decrease of one day.

[13] By combining data from the Federal Statistical Office on the total number of employees obliged to pay social insurance contributions in the different years and age groups with the SOEP data, we check the plausibility and sensitivity of this estimate. Using this method, we also control for panel attrition. To calculate the two effects, we multiply the official employment data by SOEP absence rates and income data and get a similar estimate of $(2.21 + 1.98)/2 = €2.1$ billion per year (German Federal Statistical Office, 1996, 1998).

[14] Both figures also include benefits for civil servants; however, since there was no change in sick pay regulations for civil servants, this is likely to cancel out.

costs and job losses by means of general macroeconomic equilibrium models (Zika, 1997; Feil et al., 2008; Meinhardt and Zwiener, 2005). If we relate the estimated increase in labor costs to the findings of these studies simply using the rule of proportion, we would obtain reform-induced job losses in the range of 40,000 to 80,000.[15]

However, in Germany, dismissal protection is among the strictest worldwide. The very inflexible German labor market might have triggered other attempts at compensation as well. We provide empirical evidence on how employers might have reacted to the shock to labor costs in the last four columns of Table 8. Again, we use the same OLS-DiD model as before but now we use four different outcome measures. Column (5) measures the job turnover or mobility rate. The outcome measure indicates whether respondents changed jobs between the beginning of the year prior to the interview and the interview. We find the coefficient to be insignificant and very close to zero in magnitude.

Column (6) uses the number of overtime hours per week as outcome measure. Interestingly, we find a highly significant increase in overtime of about half an hour per week. Columns (7) and (8) yield further hints as to how employers might have reacted to the positive shock to labor costs in a highly regulated labor market: by means of wage decreases relative to other occupations. Column (7) yields a highly significant relative decline in gross wages in the range of about €135 per month for all private-sector employees. In addition, in column (8), we find a smaller and imprecisely estimated wage decline for newly hired employees.

In this context, it is important to know that, in Germany, there is strong tradition of autonomy in collective bargaining, which is also referred to as "Bismarckian corporatism." This means that the wage level and most other work conditions such as overtime compensation or fringe benefits are solely subject to negotiations between unions and employer's representatives. Politicians usually do not implement laws that target these fields.[16] While we do not claim that the relative increase in overtime and the relative decrease in wages can be unambiguously traced back to the increase in absence rates and labor costs, we argue that it is at least highly likely that substantial parts of these effects were triggered by the reform. As such, we have provided empirical evidence on how work conditions in a highly regulated labor market might be adjusted in equilibrium as a reaction to an increased obligation for employers to provide social insurance benefits.

[15] As compared to a working population of about 35 million. In this very rough calculation, we completely ignore any other (general equilibrium) effects that might have been triggered by the reform.

[16] We have not found any laws that affected overtime or wages directly and were implemented in the period under consideration. However, the new center-left coalition tightened dismissal protection legislation, which might have indirectly affected these parameters.

6 Conclusion

This article empirically studied the effects of increasing the level of statutory sickness insurance benefits in Germany. The findings illustrate how social insurance interacts with a labor market that is characterized by Bismarckian corporatism. We show that an increase in statutory sick pay causally led to a decrease in employee attendance. We also provide evidence on the underlying mechanisms and of heterogeneity in the reform effects. Moreover, since (short-term) sick pay is employer-provided in Germany, we calculate the magnitude of this positive shock to labor costs and empirically study how the labor market adjusted to it.

Making good on an election campaign promise, the new center-left coalition government increased statutory short-term sick pay for private-sector employees in Germany from 80 to 100 percent of foregone gross wages as of January 1, 1999. As a result, employers were required to provide sick pay for up to six weeks per illness, without any additional benefit caps. Public-sector employees, apprentices, and the self-employed were not affected by the benefit increase.

The first part of the paper showed, based on rich SOEP panel data, how increasing insurance coverage causally affected the sick leave behavior of employees. Our identification strategy made use of conventional parametric difference-in-differences models but also non-parametric and combined approaches to check the robustness of the results. Moreover, the panel data structure allowed us to eliminate or avoid the typical pitfalls of evaluation studies, such as selection effects and sample attrition. Our findings suggest that the reform of the statutory sick pay scheme has led to an increase in sick leave of about one day per year and employee among the reform's target group. This represents an increase of about ten percent. Allowing for imperfect compliance at the individual level, alternative estimates show that the increase in sick pay was associated with a 20 percent increase in annual absence rates. Hence, the implied arc elasticity of total absence days with respect to the replacement level is 0.9.

The second part of the paper shed more light on the underlying mechanisms. We found a great amount of heterogeneity in response behavior to the policy reform and evidence that health status was the key driver of these behavioral reactions. It was primarily employees in poor health who made increased use of sick leave. While this finding is in line with the notion that a decrease in presenteeism was mainly responsible for the moral hazard effect, another finding is more consistent with a shirking explanation: we do not find any evidence that the increase in sick leave coverage improved employee health.

Finally, we provide empirical evidence as to how employers may have reacted to the increase in statutory benefits. Our calculations suggest that labor costs increased by about €1.8 billion per year due to the reform. This figure is in line with official data. Applied to the findings of other studies that were conducted based on general equilibrium models for Germany at that time, this increase in labor costs would translate into job losses of between 40,000 and 80,000. However, due to the strict dismissal protection in Germany, employers might have tried to adjust to the new labor market conditions in other ways. Indeed, we obtain empirical evidence suggesting that overtime hours increased and wages decreased in the private sector relative to other occupational groups in the aftermath of the reform.

All in all, this study provides detailed empirical evidence on how sickness absence insurance functions. Moreover, it shows how social insurance systems are linked to the labor market and what mechanisms might be triggered when exogenously increasing social insurance benefits in a regulated labor market. In this respect, the article also contributes to the debate in the US about the effects of implementing universal statutory sick leave on the federal level. The policy relevance of this topic is reflected in the Healthy Families Act' currently introduced before both houses of Congress.

References

Abadie, A., D. Drukker, J. L. Herr, and G. W. Imbens (2004). Implementing matching estimators for average treatment effects in Stata. *Stata Journal 4*(3), 290–311.

Abadie, A. and G. W. Imbens (2007). Bias corrected matching estimators for average treatment effects. Working paper, Havard University. http://ksghome.harvard.edu/ aabadie/bcm.pdf, last accessed at June 29, 2009.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion* (1 ed.). Princeton University Press.

Askildsen, J. E., E. Bratberg, and Ø. A. Nilsen (2005). Unemployment, labor force composition and sickness absence: A panel study. *Health Economics 14*, 1087–1101.

Bertrand, M., E. Duflo, and M. Sendhil (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics 119*(1), 249–275.

Besley, T. and A. Case (2000). Unnatural experiments? Estimating the incidence of endogenous policies. *Economic Journal 110*(467), 672–694.

Bound, J. (1989). The health and earnings of rejected Disability Insurance applicants. *American Economic Review 79*(3), 482–503.

Burkhauser, R. V., J. S. Butler, and G. Gumus (2004). Dynamic programming model estimates of Social Security Disability Insurance application timing. *Journal of Applied Econometrics 19*(6), 671–685.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications* (1 ed.). Cambridge University Press.

Campolieti, M. (2004). Disability Insurance benefits and labor supply: some additional evidence. *Journal of Labor Economics 22*(4), 863–890.

Chandra, A. and A. A. Samwick (2005). Disability risk and the value of Disability Insurance. NBER working papers, National Bureau of Economic Research. www.nber.or, last accessed at September 29, 2010.

Chen, S. and W. van der Klaauw (2008). The work disincentive effects of the Disability Insurance program in the 1990s. *Journal of Econometrics 142*(2), 757–784.

Cochran, W. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics 24*(2), 295–313.

Curington, W. P. (1994). Compensation for permanent impairment and the duration of work absence: evidence from four natural experiments. *The Journal of Human Resources 29*(3), 888–910.

Donald, S. G. and K. Lang (2007). Inference with difference-in-differences and other panel data. *The Review of Economics and Statistics 82*(2), 221–233.

Feil, M., S. Klinger, and G. Zika (2008). Der Beschäftigungseffekt geringerer Sozialabgaben in Deutschland: Wie beeinflusst die Wahl des Simulationsmodells das Ergebnis? *Journal of Applied Social Science (Schmollers Jahrbuch) 128*(3), 431–460.

Fischer, G., F. Janik, D. Mller, and A. Schmucker (2008). The iab establishment panel–from sample to survey to projection. FDZ Methodenreport 01/2008, The Research Data Centre (FDZ) of the Federal Employment Service in the Institute for Employment Research (IAB). $http://doku.iab.de/fdz/reporte/2008/MR_01-08_en.pdf$, last accessed at September 30, 2010.

German Federal Statistical Office (1996). *Statistical Yearbook 1996 for the Federal Republic of Germany*. Metzler-Poeschel.

German Federal Statistical Office (1998). *Statistical Yearbook 1998 for the Federal Republic of Germany*. Metzler-Poeschel.

German Federal Statistical Office (1999). *Statistical Yearbook 1999 for the Federal Republic of Germany*. Metzler-Poeschel.

German Federal Statistical Office (2001). *Statistical Yearbook 2001 for the Federal Republic of Germany*. Metzler-Poeschel.

German Federal Statistical Office (2009). *Labour market: registered unemployed*. www.destatis.de, last accessed at December 3, 2009.

Gruber, J. (2000). Disability Insurance benefits and labor supply. *Journal of Political Economy 108*(6), 1162–1183.

Heckman, J. J., H. Ichimura, and P. Todd (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies 65*(2), 261–94.

Henrekson, M. and M. Persson (2004). The effects on sick leave of changes in the sickness insurance system. *Journal of Labor Economics 22*(1), 87–113.

Ichino, A. and G. Maggi (2000). Work environment and individual background: explaining regional shirking differentials in a large italian firm. *The Quarterly Journal of Economics 115*(3), 1057–1090.

Ichino, A. and E. Moretti (2009). Biological gender differences, absenteeism, and the earnings gap. *American Economic Journal: Applied Economics 1*(1), 183–218.

Imbens, G. W. (2008). The evaluation of social programs: some practical advice. Presentation, 2^{nd} IZA/IFAU Conference on Labour Market Policy Evaluation. October 11, 2008.

Imbens, G. W. and D. B. Rubin (2009). *Causal Inference in Statistics and the Social Sciences* (1 ed.). Cambridge and New York: Cambridge University Press. forthcoming.

Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Jahn, J. (1998). Lohnfortzahlung: Gerichte stehen vor Herkulesaufgabe. *Handelsblatt 124: 02.07.1998*, 4.

Johansson, P. and M. Palme (1996). Do economic incentives affect work absence? Empirical evidence using Swedish micro data. *Journal of Public Economics 59*(1), 195–218.

Johansson, P. and M. Palme (2002). Assessing the effect of public policy on worker absenteeism. *Journal of Human Resources 37*(2), 381–409.

Johansson, P. and M. Palme (2005). Moral hazard and sickness insurance. *Journal of Public Economics 89*, 1879–1890.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review 76*(4), 604–620.

Lechner, M. (2002). Program heterogeneity and propensity score matching: an application to the evaluation of active labour market policies. *The Review of Economics and Statistics 84*(2), 205–220.

Medizinischer Dienst der Krankenversicherung (MDK) (2008). www.mdk.de, last accessed at June 25, 2009.

Meinhardt, V. and R. Zwiener (2005). Gesamtwirtschaftliche Wirkungen einer Steuerfinanzierung versicherungsfremder Leistungen in der Sozialversicherung. Politikberatung kompakt 7, German Institute for Economic Research (DIW) Berlin. http://www.diw.de, last accessed at December 19, 2008.

Meyer, B. D., W. K. Viscusi, and D. L. Durbin (1995). Workers' compensation and injury duration: evidence from a natural experiment. *American Economic Review 85*(3), 322–340.

Puhani, P. A. and K. Sonderhof (2010). The effects of a sick pay reform on absence and on health-related outcomes. *The Journal of Health Economics 29*(2), 285–302.

Ridinger, R. (1997). Einfluss arbeitsrechtlicher Regelungen auf die Beschäftigungsentwicklung im Handwerk–Ergebnisse von Befragungen von Handwerksbetrieben im 3. Quartal 1997. Dokumentation, Zentralverband des Deutschen Handwerks. http://www.zdh.de, last accessed at June 19, 2009.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Rosenbaum, P. R. and D. B. Rubin (1984). Reducing the bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association 79*(387), 516–524.

Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (1998). *Vor weitreichenden Entscheidungen*. Metzler-Poeschel.

Social Security Administration (2006). *Annual Statistical Supplement 2006, Table 9.A2*. http://www.ssa.gov/policy/docs/statcomps/supplement/2006/9a.html, last accessed at March 19, 2009.

Social Security Administration (2008). *Annual Statistical Supplement 2006, Table 9.C1*. http://www.ssa.gov/policy/docs/statcomps/supplement/2008/9c.html, last accessed at March 19, 2009.

Wagner, G. G., J. R. Frick, and J. Schupp (2007). The German Socio-Economic Panel Study (SOEP) - evolution, scope and enhancements. *Journal of Applied Social Science (Schmollers Jahrbuch) 127*(1), 139–169.

Ziebarth, N. R. and M. Karlsson (2010). A natural experiment on sick pay cuts, sickness absence, and labor costs. *Journal of Public Economics* (forthcoming).

Zika, G. (1997). Die Senkung der Sozialversicherungsbeiträge. IAB Werkstattbericht 7, Research Insitute of the Federal Employment Agency (IAB).

Figure 1: Distribution of Propensity Scores Showing Region of Common Support

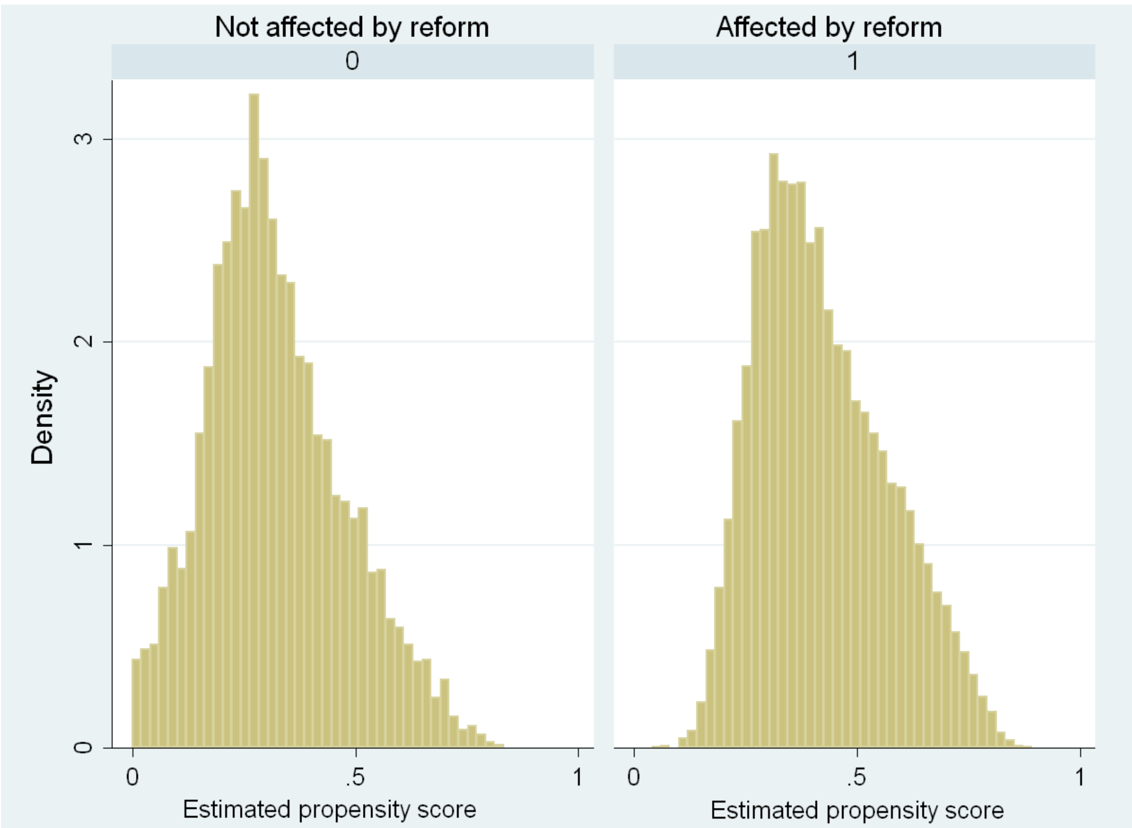

Figure 2: Average Sickness Absence Days for Treatment and Control Group over Time

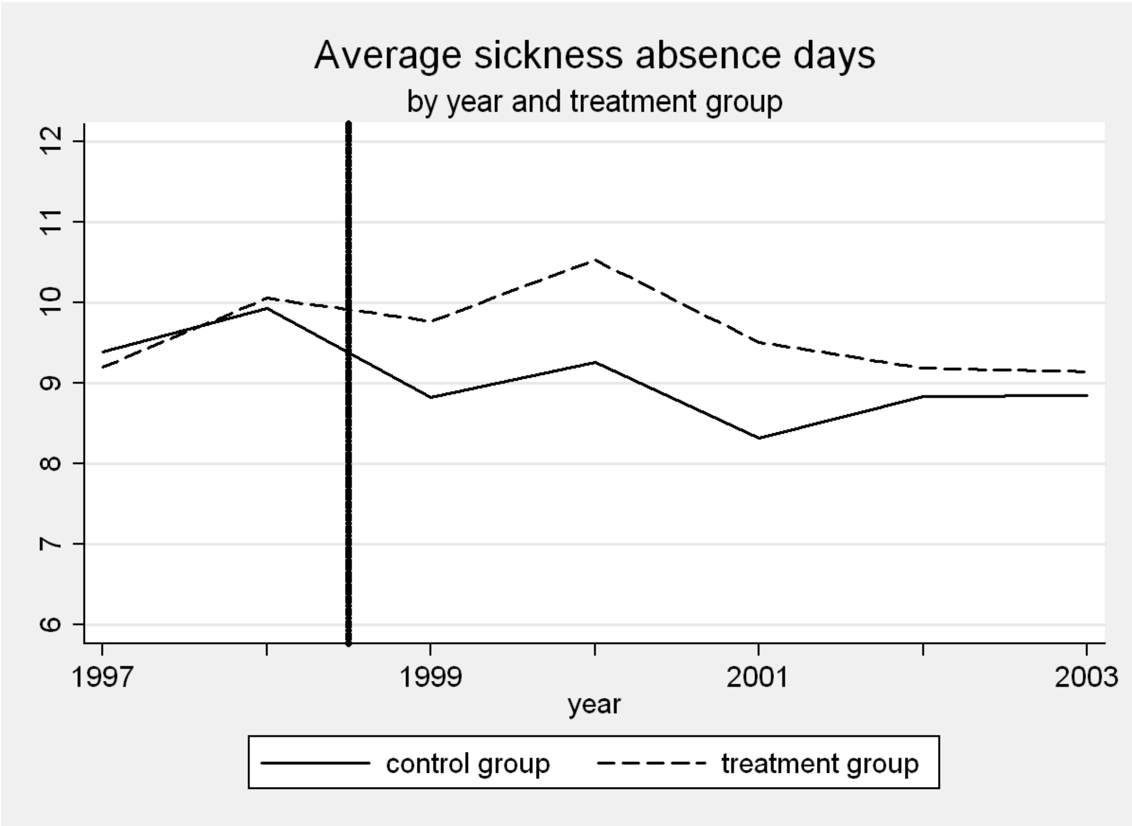

Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Treatment Group	0.6566	0.4749	0	1	23,058
Daysabs	9.7793	24.9027	0	365	23,058
A: Variables used in main specifications					
Personal characteristics					
Female	0.3188	0.466	0	1	23,058
Age	39.69	8.22	25	55	23,058
Age squared	1,643	661	625	3,025	23,058
Immigrant	0.1433	0.3504	0	1	23,058
East German	0.2654	0.4415	0	1	23,058
Partner	0.7944	0.4042	0	1	23,058
Married	0.6803	0.4664	0	1	23,058
Children	0.4795	0.4996	0	1	23,058
Disabled	0.0438	0.2048	0	1	23,058
Health good (best two of five SAH categories)	0.6355	0.4813	0	1	23,058
Health bad (worst two of five SAH categories)	0.0811	0.2731	0	1	23,058
No sports	0.3776	0.4848	0	1	23,058
Educational characteristics					
Drop out	0.0269	0.1618	0	1	23058
Degree after 8 years' schooling	0.2883	0.453	0	1	23,058
Degree after 10 years' schooling	0.3659	0.4817	0	1	23,058
Degree after 12 years' schooling	0.0512	0.2204	0	1	23,058
Degree after 13 years' schooling	0.1929	0.3946	0	1	23,058
Other degree	0.0746	0.2627	0	1	23,058

Continued on next page...

... Table 1 continued

Variable	Mean	Std. Dev.	Min.	Max.	N
Years with company	9.5069	8.4721	0	41	23,058
Trained for job	0.6063	0.4886	0	1	23,058
Job characteristics					
New job	0.1573	0.3641	0	1	23,058
Blue-collar worker	0.3707	0.483	0	1	23,058
White-collar worker	0.4602	0.4984	0	1	23,058
One man company	0.0328	0.1781	0	1	23,058
Small size company	0.2441	0.4296	0	1	23,058
Medium size company	0.2754	0.4467	0	1	23,058
Large company	0.2211	0.415	0	1	23058
Very large company	0.2266	0.4186	0	1	23058
Gross wage per month	2,467	1,278	404	28,632	23,058
Annual state unemployment rate	11.36	4.53	5.4	21.7	23,058
B: Variables used in extended analyses					
Long-term absence	0.0558	0.2296	0	1	23,058
Job loss	0.1102	0.3131	0	1	23,058
Not impaired by health	0.7826	0.4125	0	1	23,058
Severely impaired by health	0.0236	0.1518	0	1	23,058
Low health satisfaction (0-4 on scale up to 10)	0.0528	0.2237	0	1	23,058
High health satisfaction (10 on scale up to 10)	0.0528	0.2237	0	1	23,058
Overtime hours per week	2.6811	3.8891	0	23.1	20,732
Gross wage last month (job change previous year)	2,170	1,135	409	12,782	3627
Low job satisfaction (0-4 on scale up to 10)	0.055	0.2281	0	1	22,347
Very worried about job security	0.1566	0.3634	0	1	22,503
Job makes no fun ('97; no job changers)	0.1301	0.3365	0	1	12,786
Not religious ('97)	0.3649	0.4814	0	1	15,537
Sickness should be insured by state ('97)	0.3804	0.4855	0	1	15,537
Sickness should be insured privately('97)	0.0857	0.2799	0	1	15,537
Expects job loss within 2 years ('98; no job changers)	0.0875	0.2826	0	1	16,771
Expects promotion within 2 years ('98; no job changers)	0.1958	0.3968	0	1	16,771
Firm reduced workforce last year ('99; no job changers)	0.2795	0.4488	0	1	17,201
Control life ('99)	0.2888	0.4532	0	1	17,351
Can influence life ('99)	0.4083	0.4915	0	1	17,351
Need to work hard for success ('99)	0.5257	0.4994	0	1	17,351
No work council in firm ('01)	0.3841	0.4864	0	1	16,161

Variables with years in parenthesis were only surveyed in the corresponding year. When the information sampled refers to the workplace, only respondents who still work at the same workplace are kept. For example, respondents who answered the work council question in 2001 are kept in all years in which they were interviewed and worked at the same workplace as in 2001. For variables that were only surveyed in one year but do not contain workplace information, we keep the respondents in all years in which they were interviewed and assume time invariance. For example, respondents who in 1999 stated that one would need to work hard for success are kept in all years in which they were interviewed. It is assumed that they did not change their attitude over time.

Table 2: Sample Means of Treatment and Control Group: Raw, Matched, and Blocked Sample

	Raw Sample			***Matched Sample***			***Blocked Sample***		
Covariates	**Treat. group**	**Control group**	**Norm. diff.**	**Treat. group**	**Control group**	**Norm. diff.**	**Treat. group**	**Control group**	**Norm. diff.**
Age	39.07	40.9	0.157	39.182	40.614	0.125	39.07	39.35	0.082
Female	0.277	0.398	0.182	0.301	0.374	0.110	0.277	0.424	0.134
Partner	0.798	0.788	0.017	0.792	0.791	0.002	0.798	0.781	0.060
Married	0.676	0.688	0.018	0.676	0.685	0.014	0.676	0.675	0.058
Immigrant	0.178	0.077	0.217	0.151	0.085	0.147	0.178	0.079	0.086
Children	0.489	0.462	0.037	0.480	0.463	0.024	0.489	0.433	0.073
Disabled	0.042	0.048	0.020	0.043	0.047	0.015	0.042	0.042	0.046
Health good	0.633	0.641	0.012	0.632	0.639	0.010	0.633	0.609	0.030
Health bad	0.081	0.082	0.004	0.080	0.083	0.007	0.081	0.079	0.036
8 years of schooling	0.326	0.216	0.178	0.309	0.237	0.115	0.326	0.217	0.046
10 years of schooling	0.349	0.397	0.070	0.364	0.418	0.078	0.349	0.409	0.082
13 years of schooling	0.151	0.273	0.213	0.168	0.224	0.100	0.151	0.230	0.060
Trained for job	0.565	0.685	0.175	0.584	0.662	0.114	0.565	0.670	0.057
New job	0.177	0.119	0.117	0.165	0.124	0.082	0.177	0.115	0.042
Years with company	8.748	10.958	0.184	9.093	10.563	0.122	8.748	11.576	0.087
White collar	0.495	0.394	0.144	0.482	0.430	0.075	0.495	0.492	0.195
Gross wage/1,000	2,392	2,611	0.117	2,408	2,496	0.052	2,392	2,520	0.084
State unemployment rate	11.148	11.751	0.094	11.301	11.792	0.076	11.148	11.830	0.106

"Norm. diff." stands for "Normalized difference" which is calculated according to $\frac{\bar{s}_1-\bar{s}_0}{\sqrt{\sigma_1^2+\sigma_0^2}}$, where $\bar{s}_1$ is the sample mean of the covariate for the treatment group and $\bar{\sigma}_0^2$ stands for the variance of the covariate within the control group. The "matched sample" has been generated by means of five-to-one nearest neighbors matching based on the propensity score. "Blocked sample" means that the sample was blocked to guarantee identicial propensity scores within blocks. Here, the propensitiy score is the probability of belonging to the treatment group and was estimated by a logit model under the inclusion of the displayed covariates and *(years with company)*^{2}, *(years with company)*× *female*, *(years with company)*×*(trained for job)*, *(annual state unemployment rate)*^{2}, *(gross wage)*^{2}, *(gross wage)*×*female*, *(gross wage)*×*(white collar)*, *(gross wage)*×*(13 years of schooling)*, *(gross wage)*×*married*. The covariates in levels as well as their interactions to estimate the propensity score were selected according to likelihood ratio tests on zero coefficients as decribed in Imbens (2008). After the PS estimation, in the blocked sample, 221 observations (0.01%) are not considered since they lie outside the common support which is [0.0341; 0.9956]. The number of blocks is twelve; the smallest block contains 46 respondents in the treatment and 317 respondents in the control group. In total, the raw sample contains 23,058 observations, the matched sample contains 19,190 observations, and the blocked sample contains 22,837 observations.

Table 3: Difference-in-Differences Estimation: Parametric, Non-Parametric, and Combined Methods

	Regression		*Matching*		*Matching + Regression*	
Variable	**OLS**	**ZINB-2**	**blocking**	**nearest neighbors**	**blocking + regression**	**n.neighbors + regression**
DiD	1.3659*	1.0181**	1.1382***	1.1203***	0.8973**	0.9986***
	(0.7097)	(0.4684)	(0.4064)	(0.4577)	(0.4223)	(0.3307)
Covariates employed						
Job characteristics	yes	yes	yes	yes	yes	yes
Educational characteristics	yes	yes	yes	yes	yes	yes
Personal characteristics	yes	yes	yes	yes	yes	yes
Regional unemployment rate	yes	yes	yes	yes	yes	yes
Time dummies	yes	yes	no	no	yes	no
State dummies	yes	yes	no	no	yes	no
N	23,058	23,058	19,071	19,040	22,837	23,058

* $p<0.1$, ** $p<0.05$, *** $p<0.01$; standard errors are in parentheses. In the parametric specifications, they are adjusted for intrapersonal correlations. The estimate in column (2) is the marginal effect, calculated at the means of the covariates except for the post reform dummy (=1), the treatment group dummy (=1), the year 1999 dummy (=1), the year 2000 dummy (=1), and *DiD* (=1). ZINB-2 stands for *Zero-Inflated Negative Binominal Model 2.* In columns (3) and (4), the "matched sample" of Table 2 is the underlying sample. In column (3), the propensity score (PS) of belonging to the treatment group in post-reform years (*DiD*=1) is estimated, based on a logit model and the same covariates as in Table 2. Based on this PS, the sample is stratified into eleven blocks, each with an equal PS for treated and non-treated. 119 observations (12 treated) are outside the region of common support. Then, the block-specific treatment effects—the difference in average outcomes for treated (*DiD*=1) and non-treated (*DiD*=0)—are weighted by the number of treated to obtain the overall average treatment effect on the treated. In column (4), the average treatment effect on the treated is obtained by five-to-one nearest neighbors matching. In that specification, 150 observations lie outside the common support (31 treated). Standard errors in column (4) are obtained by bootstrapping with 100 replications. In column (5), the "blocked sample" of Table 2 is used. Then within each block, a ZINB-2-DiD regression is performed. Finally, the within block estimates are weighted by the number of treated observations to obtain the overall treatment effect on the treated. In column (6), five-to-one nearest neighbors matching and regression are combined. As explained in Abadie et al. (2004), the estimator is bias corrected and allows for heteroskedastic errors. In all columns, except for columns (3) and (4), the number of treated observations is 7,199. In column (3) [(4)], the number of treated observations is 7,187 [7,168] since 12 [31] observations lie outside the region of common support.

Table 4: Difference-in-Differences Estimation: Specific Samples

Variable	*Full Sample* Intended	*Full Sample* Implemented	*Full Sample* Implemented clustered higher	*Private Sector Employees* Intended	*Private Sector Employees* Implemented	*Private Sector Employees* Implemented clustered higher
DiD	1.8077*	2.1282**	2.1282***	1.4909	2.5496	2.5496**
	(0.9345)	(1.0838)	(0.0982)	(1.2834)	(4.6470)	(0.1571)
Covariates employed						
Job characteristics	yes	yes	yes	yes	yes	yes
Educational characteristics	yes	yes	yes	yes	yes	yes
Personal characteristics	yes	yes	yes	yes	yes	yes
Regional unemployment rate	yes	yes	yes	yes	yes	yes
Time dummies	yes	yes	no	no	yes	no
State dummies	yes	yes	no	no	yes	no
N	11,929	11,929	11,929	4,921	4,921	4,921

* p<0.1, ** p<0.05, *** p<0.01; standard errors are in parentheses. All six models are based on a detailed review of all collective agreements in the main industries. Instead of assigning a one to every respondent in the private sector as in the previous table, we now use the fraction of workers receiving 100 percent sick pay. In addition to public sector employees and the self-employed, i.e., *Treatment Group=0*, who did not experience changes in their sick pay throughout the entire period under consideration, one can differentiate between the following subgroups: a.) industries that unambiguously codify in their collective agreements that solely the statutory sick pay minimum is provided, e.g., the construction and agriculture sector, b.) industries that unambiguously codify in their collective agreements that 100 percent sick pay is provided, e.g., the chemical industry or credit and insurance industries. Firms that are not covered by collective agreements are very likely to only provide sick pay according to the statutory minimum standards. For the period under consideration, the SOEP includes information on the industries of the respondents but not on whether the employer is covered by a collective agreement. Hence, we assign repondents in columns (2), (3), (5), and (6) according to the following criteria: the treatment indicator equals zero for every respondent who worked in an industry that strictly paid statutory minimum sick leave in the pre-reform period (within this group, employees covered by collective agreements as well as employees not covered by collective agreements experienced the increase in sick pay). In addition, we use information from the IAB Establishment Panel (Fischer et al., 2008) on the share of collective bargaining coverage in private sector industries that provided 100 percent sick pay. Hence, we exclude all private sector industries for which the degree of implementation is unknown. This is why the number of observations is smaller than in the tables above. In contrast to columns (4) to (6), columns (1) to (3) include public sector employees and the self-employed, who are codified as being fully treated throughout. Columns (4) to (6) solely rely on private sector employees. Columns (1) and (4) are for the purpose of comparison and assign respondents to the treatment and control group using the same criteria as in the previous analysis. Columns (1), (2), (4), and (5) do not cluster standard errors. Columns (3) and (6) cluster standard errors on the (treatment indicator)×year level.

Table 5: Robustness Checks

Model	+ lagged daysabs	observed pre- and post	no job changers	no post-reform industry changers	clustered at state × year level	clustered at industry × year level	impact on long-term absenteeism
OLS	1.5676**	1.1396	1.3709*	1.1918*	1.3659**	1.3659**	-0.0034
	(0.7149)	(0.7281)	(0.7794)	(0.7156)	(0.5992)	(0.6857)	(0.0064)
N	19,223	15,115	19,431	21,478	23,058	23,058	23,058

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers, except for column (5) where they are clustered on state×year (64 cluster) and column (6) where they are clustered on the industry × year (242 cluster) level. All specifications are as in column (1) of Table 3 except for the following: the model in column (1) contains the lagged number of annual absence days as an additional covariate. The model in column (2) includes only those who were observed at least once (working) in the pre-reform years *and* the post-reform years. The model in column (3) excludes all those who have changed their jobs in the year prior to the interview. The model in column (4) excludes private sector employees who have changed their industry branch in the post-reform period. The model in in column (7) estimates the reform effect on the incidence of long-term absenteeism, i.e., a sickness period of more than six weeks.

Table 6: Difference-in-Differences Estimation on the Number of Absence Days: Placebo Estimates

Model	**2000**	**2001**
OLS	0.2750	-0.6487
	(0.5980)	(0.6073)
ZINB-2	0.1925	-0.1220
	(0.4485)	(0.4399)
nearest neighbors	0.3468	-0.0617
+ regression	(0.4459)	(0.4129)
N	25,692	27,912

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. Both columns make use of two pseudo pre- and two pseudo post-reform years, i.e., column (1) includes the waves 1999-2002 and column (2) includes the waves 2000-2003. Marginal effects for the ZINB-2 are calculated at the means of the covariates except for the post reform dummy (=1), the treatment group dummy (=1), the year dummies (=1 or =0), and *DiD* (=1). Every cell stands for one model.

Table 7: Assessing Heterogeneity in Reform Effects

Panel A: Personal characteris						
	female	**over 40**	**gross wage > median**	**partner**	**job loss previous year**	**highest school degree**
DiD × [column]	-1.0517	-0.1118	-0.8639	1.6746*	-0.3316	-0.0567
	(0.7982)	(0.7395)	(0.7358)	(0.8962)	(1.3639)	(0.7339)
DiD	1.6458**	1.4156*	1.7914**	0.0257	1.4406**	1.3746*
	(0.7396)	(0.7383)	(0.8172)	(1.0230)	(0.7272)	(0.7248)
Covariate [column]	0.7601	-0.0683	0.0175	-1.4610**	5.1830***	-5.6689***
	(1.1507)	(0.7071)	(0.5717)	(0.7187)	(0.9158)	(1.7624)
Panel B: Health status						
	health very good	**health bad or poor**	**high health satisfaction**	**low health satisfaction**	**not impaired by health**	**disabled**
DiD × [column]	-0.0588	3.3827	-1.2153*	5.7636	-1.1841	9.4904**
	(0.6087)	(2.7622)	(0.7084)	(7.1318)	(1.2762)	(4.1585)
DiD	1.4670**	1.0919	1.4791**	1.1160	2.2897	0.9503
	(0.7279)	(0.6963)	(0.7217)	(0.7684)	(1.3583)	(0.6919)
Covariate [column]	-2.9769***	14.7947***	-1.3153***	10.0281***	-4.801***	7.2174***
	(0.3881)	(1.6259)	(0.4768)	(1.9772)	(0.7618)	(1.9387)

* $p<0.1$, ** $p<0.05$, *** $p<0.01$; standard errors in parentheses are adjusted for clustering on person identifiers. All specifications estimate the model in equation (1) by OLS. Additionally, all models include an interaction term between *DiD* and the corresponding covariate in the column header. *Female* is a dummy variable with a one for females. *Over 40* is a dummy variable with a one for respondents over the age of 40. *Gross wage > median* is a dummy variable with a one for respondents who earn more than €2,199 per month. *Partner* has a one for respondents in a partnership. *Job loss previous year* is a dummy variable that indicates whether the employee changed the job in the previous year. *Highest school degree* means holding a certificate after 13 years of schooling. In Panel B, the first two columns make use of dummy variables that were generated from self-assessed health (SAH). *Health very good* has a one for respondents who indicated to have the best health status on the five-category SAH scale. *Health bad or poor* has a one for respondents who rated themselves in the worst two SAH categories. *Low health satisfaction* are the collapsed lowest four categories on an eleven-category scale on health satisfaction. *High health satisfaction* has a one for those ranked in the best health satisfaction category. *Not impaired by health* is generated from the answer category "Not at all" to the following question: "Aside from minor illnesses, does your health prevent you from completing everyday tasks like work around the house, paid work, studies, etc.? To what extent?" *Disabled* has a one for respondents who are officially certified as disabled. All models have 23,068 observations. The descriptive statistics for all column-header variables used are shown in Table 1.

Panel C: Subjective workplace characteristics						
	low job satisfaction	**very worried (job security)**	**job makes no fun**	**job loss likely within 2 yrs. ('98)**	**promotion likely within 2 yrs. ('98)**	
DiD × [column]	1.3064	-0.6709	0.2375	-0.8603	-0.4035	
	(2.4201)	(1.1887)	(2.061)	(1.8713)	(0.9446)	
DiD	1.2119*	1.5983**	1.7791*	1.9334**	1.8802**	
	(0.6796)	(0.7018)	(0.9991)	(0.7919)	(0.8332)	
Covariate [column]	4.0110***	1.2966*	1.1361	2.4224**	-0.4657	
	(1.2911)	(0.7185)	(0.9922)	(1.1416)	(0.5025)	

Panel D: Personality traits and attitudes						
	not religious	**sickness should be insured by state**	**sickness should be insured privately**	**can influence life**	**control life 1999**	**need to work hard for success**
DiD × [column]	-0.7353	0.1606	3.4534	2.0955**	2.3197**	1.6421*
	(1.0488)	(1.0049)	(2.2170)	(0.4868)	(1.0413)	(0.8706)
DiD	1.8233**	1.4816	1.2793	0.9006	1.1224	0.9133
	(0.9419)	(0.9153)	(0.8655)	(0.8803)	(0.8519)	(08893)
Covariate [column]	1.5459**	-0.0942	0.1208	-0.7349	-0.3707	0.2777
	(0.6958)	(0.5712)	(0.1650)	(0.4868)	(0.5403)	(0.5014)

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. All specifications estimate the model in equation (1) by OLS. Additionally, all models include an interaction term between *DiD* and the corresponding covariate in the column header. For those variables that were only sampled in one specific year and that relate to the workplace (Panel C), we keep only respondents in years in which they worked at the sample workplace as in the correponding year. For those variables that were only asked in one specific year and represent personality traits or attitudes (Panel D), we keep the respondents in all years in which they answered the SOEP questionnaire. In both cases, time persistence is assumed. For example, respondents who answered (only) in 1997 that their job would make no fun are kept in all years in which they had the same workplace as in 1997. *Low job satisfaction* in column (1) of Panel C stands for the lowest four categories on an eleven-category scale on job satisfaction (22,347 obs.) and *very worried about job security* has a one for respondents who answered "very concerned" to the following question: "Are you concerned about your job security" (22,503 obs.). *Job makes no fun* was only sampled in 1997 and has a one for those who answered "applies completely" or "applies more or less" towards the statement "I do not enjoy my work." (12,783 obs.). In 1998, respondents were asked whether they believe that they would loose their job (get promoted) within the next two years. Those who answered "very likely" or "likely" are represented by the dummy variables that are used in columns (4) and (5) of Panel C (16,771). The variables used in columns (1) to (3) of Panel D were only sampled in 1997 (15,337 obs.). The variables used in columns (4) to (6) of Panel D were only sampled in 1999 (17,351 obs.) *Not religious* is a dummy variable with a one for everyone who answered "never" to the question "How often do you go to church or religious institutions?". *Sickness should be insured by the state (privately)* has a one for those who claimed that sickness should be "only" or "mostly" insured by the state (privately). *Can influence life* has a one for respondents who said that they can totally agree with the statement: "How life proceeds, depends on me." *Control life* has a one for respondents who said that they totally disagree with the statement: "I often experience that others have control over my life." *Need to work hard for success* has a one for respondents who said that they can totally agree with the statement: "One has to work hard to achieve success." The descriptive statistics for all column header variables used are shown in Table 1.

Table 8: Reform Effect on Employees' Health Status and Employers' Behavior

	Employees: health & workplace climate				***Employers: dismissals, overtime, & wages***			
	health bad or poor	**sev. impaired by health**	**low health satisfaction**	**low job satisfaction**	**job loss prev. year**	**overtime (hours/week)**	**gross wage (per month)**	**gross wage new job**
DiD	0.0024	-0.0009	-0.0001	0.0049	0.0063	0.5369***	-135.39***	-56.99
	(0.0077)	(0.0044)	(0.0066)	(0.0067)	(0.0088)	(0.1081)	(28.57)	(93.03)
Job controls	yes	yes	yes	yes	yes	yes	yes	yes
Edu controls	yes	yes	yes	yes	yes	yes	yes	yes
Personal controls	yes	yes	yes	yes	yes	yes	yes	yes
Reg. unempl. rate	yes	yes	yes	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes	yes	yes

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. All specifications estimate the model in equation (1) by OLS but use the corresponding variable in the column header as outcome measure. All outcome variables used are detailed in Table 1. The model in the last column estimates the effect on gross wages for employees who claimed to have changed their jobs in the year prior to the interview (3,759 obs.). The models in the first three columns have 23,058 observations, the model in column (4) has 22,347 observations, and the models in columns (5) to (7) have 20,732 observations.