DISCUSSION PAPER SERIES

IZA DP No. 7250

The Effects of Expanding the Generosity of the Statutory Sickness Insurance System

Nicolas R. Ziebarth
Martin Karlsson

February 2013

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Effects of Expanding the Generosity of the Statutory Sickness Insurance System

Nicolas R. Ziebarth
Cornell University
and IZA

Martin Karlsson
University of Duisburg-Essen

Discussion Paper No. 7250
February 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 7250
February 2013

ABSTRACT

The Effects of Expanding the Generosity of the Statutory Sickness Insurance System*

This article evaluates an expansion of employer-mandated sick leave from 80 to 100 percent of forgone gross wages in Germany. We employ and compare parametric difference-in-difference (DID), matching DID, and mixed approaches. Overall workplace attendance decreased by at least 10 percent or 1 day per worker per year. We show that taking partial compliance into account increases coefficient estimates. Further, heterogeneity in response behavior was of great importance. There is no evidence that the increase in sick leave improved employee health, a finding that supports a shirking explanation. Finally, we provide evidence on potential labor market adjustments to the reform.

JEL Classification: H51, I18, J22, J32

Keywords: difference-in-differences estimation, sickness absence, generosity of social insurance, employer sick leave mandate, natural experiment, SOEP

Corresponding author:

Nicolas R. Ziebarth
Cornell University
Department of Policy Analysis and Management (PAM)
106 Martha Van Rensselaer Hall
Ithaca, NY 14850
USA
E-mail: nrz2@cornell.edu

* We would like to thank anonymous referees as well as the editor, Edward Vytlacil, for extremely helpful comments and suggestions that substantially improved the quality of this paper. Moreover, we thank Daniela Andrén, Patrick Arni, Ghazala Azmat, Tim Barmby, Mattias Bokenblom, Jörg Breitung, Rich Burkhauser, Antonio Cabrales, John Cawley, Laurens Cherchye, Antonio Ciccone, Irma Clots-Figueras, Meltem Daysal, Miguel A. Delgado, Stefano DellaVigna, Vincenzo Denicolò, Per-Anders Edin, Liran Einav, Marco Francesconi, Peter Frederiksson, Christina Gathmann, Jon Gemus, Rick Geddes, Murat Genç, Albrecht Glitz, Rita Ginja, Lorenz Goette, Libertad Gonzalez, Nils Gottfries, David Granlund, Dan Hamermesh, Barbara Hanel, Lars Hultkrantz, Guido Imbens, Per Johansson, Don Kenkel, Stephen Knowles, Mathias Kifmann, Tobias Klein, Rafael Lalive, Anna Larsson, Michael Lechner, Adam Lederer, Mikael Lindahl, Mike Lovenheim, Matilde Machado, Olivier Marie, Simen Markussen, Jürgen Maurer, Raymond Montizaan, Eva Mörk, Peter Nilsson, Martin Olsson, Dorian Owen, Mårten Palme, Markus Pannenberg, Matthias Parey, Steve Pischke, Per Petterson-Lidbom, James Rockey, Knut Røed, Martin Salm, Peter Sivey, Eric Smith, Peter Skogman Thoursie, Jan C. van Ours, Frederic Vermeulen, Tarja Viitanen, Johan Vikström, Roger Wilkins, Will White, Mark Wooden, and participants 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 session on "The Economic of Sickness Absence" at the 2011 meeting of the American Economic Association (AEA 2011), the Workshop on Absenteeism and Social Insurance in Uppsala, the Econometrics and Statistics Seminar at the University of Tilburg, the Economics Seminar at the University of Augsburg, the Economics Seminar at the University of Otago, the Seminar Series of the Melbourne Institute at the University of Melbourne, the Joint Empirical Social Science (JESS) Seminar at the Institute for Social & Economic Research (ISER) at the University of Essex as well as seminar participants at HEC Lausanne, Universidad Carlos III, Universitat Pompeu Fabra, the University of Leicester, the University of Edinburgh, Stockholms Universitet, Uppsala Universitet, the University of Essex, Cornell University, and the Berlin Network of Labour Market Researchers (BeNA) for their helpful comments and discussions. 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.

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: statutory sickness insurance. While statutory sick leave is almost completely unknown in the US and Canada, federally mandated sickness insurance is an integral part of social insurance systems across Europe.

Statutory sickness insurance (SI) 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 at the state level, as well as *Disability Insurance (DI)*, which replaces income losses stemming from *permanent* work absences due to disabilities and is administered at the federal level. However, the US is the only industrialized country that does not guarantee that workers receive paid leave for work-unrelated sickness (Heymann et al., 2009). Estimates suggest that half of all US workers have no access to paid sick leave; the share is much higher among low-income employees (Boots et al., 2009). Relatively few people know that six US states and Puerto Rico, a US commonwealth, 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). Moreover, San Francisco and Washington DC passed sick leave legislation at the city level in 2006 and 2008, respectively. Effective January 1, 2012, Connecticut is the first US state explicitly mandating that employers provide paid sick leave to their employees.

Not surprisingly, in the US, a heated debate about the implementation of universal statutory sick leave on the federal level has emerged. Senator Edward Kennedy first introduced a bill called the *Healthy Families Act* in the US Congress in 2005. In 2007 and 2009, it was reintroduced and would require that every US employer with more than 15 employees 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 ("presenteeism").

Although it remains unclear to what degree the findings can be applied to the US, this paper illustrates potential effects of expanding an existing federally mandated sick leave scheme in Germany. We thereby contribute to the literature in a number of ways: First of all, we provide clear-cut evidence on how a substantial increase in federally mandated sick leave benefits has causally 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, matching, as well as combined difference-in-differences (DID) approaches. Moreover, we distinguish between (*i*) "intention-to-treat" (ITT) effects, which are relevant for policy makers; and (*ii*) average treatment effects on the treated (ATT), i.e, the response in labor supply to changes in the replacement rate as actually implemented. In addition, we attempt to shed light on the mechanisms underlying the behavioral reactions by an extensive effect heterogeneity analysis and by estimating the effects on health and well-being. Furthermore, we provide empirical evidence on how employers might have reacted to this exogenous increase in employer-mandated sick leave benefits. To our knowledge, this is the first paper to attempt such a unified analysis.

The reform to which we devote our attention in this paper received only limited attention in the literature. Ziebarth and Karlsson (2010) (*ZK2010*) analyze a previous German reform—a cut in sick pay that was introduced in 1996. Puhani and Sonderhof (2010) (*PS2010*) compare the effects of this previous cut in sick pay to the subsequent increase in sick pay in 1999; the latter being the focus of this paper. Both papers establish a negative effect of the cut in sick pay on the number of sick days used: *ZK2010* observe an increase in the proportion of individuals without any sickness days by three percentage points, corresponding to a six percent increase. *PS2010* estimate a fairly robust reduction in the number of sickness days per year: on average, individuals had two days of sickness less after the cut in sick pay. The sick day effect of *PS2010* is similar to the estimate of *ZK2010*, although the two studies use different identification strategies. *PS2010* also consider alternative outcome variables such as utilization of health care services and self-assessed health. Their main conclusion is that a reduction in the sick pay leads to a large reduction in hospital stays, but there is no discernible effect on individual health. Moreover, the two authors compare switch-on effects from the 1996 reform to switch-off effects from the 1999 reform. Their results suggest that switch-off effects are larger, but the differences are typically not statistically significant.

Although using the same dataset and analyzing the same reform, this paper differs substantially from *PS2010*. It also differs substantially from *ZK2010*, who evaluate a cut in sick pay, not an increase. In our view, it is improper to analyze the two reforms jointly, since the 1999 expansion reform that we analyze in this paper was not simply a reversal of the 1996 cut reform (see *ZK2010* and footnote 4 for more details). Moreover, the identification strategy employed by *PS2010* relies on information of whether a worker was covered by a collective bargaining contract, and this question was asked only in 1995, which means that there will be substantial aging and attrition in the panel once we reach the end of the survey period required to analyze the 1999 reform. Thus, our work offers a check of the robustness of *PS2010*'s findings to the usage of a more up-to-date dataset and different identification strategies. Concerning the latter, our estimates will be based on a comparison between private and public sector workers and ITT effects, as well as a variable capturing actual implementation at the industry level and ATT effects.[1] Apart from a different identification strategy and from comparing ITT with ATT estimates, this paper applies and tests the robustness of regression, matching, and combined DID approaches.

In addition to these contributions in applied econometrics, there are at least four main research areas in the field of SI that our findings contribute to. First, the most obvious research area comprises the assessment of changes in benefit levels on sick leave behavior. As outlined above, using survey data, *ZK10* and *PS2010* focused on this question for Germany, while, using administrative data, a handful of other studies analyze this question for Scandinavian countries (see Section 2 for more details). This paper confirms previous results and

[1] We do not use the 1995 information on collective bargaining agreements used by *PS2010*. In order to use the information from 1995, *PS2010* need to drop all individuals who changed jobs in the 1996-1998 period. However, these individuals are dropped only in the relevant years, which leads to a discrepancy between the populations used in the pre- and post-reform periods. Also the sample size suffers from the requirement that data on coverage of collective bargaining agreement is only available for 1995: *PS2010* attain a sample size of 13,500, while ours, covering a shorter time period, is 23,000. Finally, *PS2010* pool private *and* public sector employees and establish treatment status using the collective bargaining question. However, as Figure 1 in *ZK2010* shows, this approach mixes treated and non-treated in the treatment and the control group since public sector employees where not affected at all, and private sector employees with collective agreements in specific industries were treated.

shows that employees adjust their sick leave behavior to changes in benefit levels as standard economic theory would predict. However, we differentiate between ITT and ATT effects and find that the latter are about twice as large. This finding is in line with the political economy of the reform, according to which only 50 percent of all employees were actually affected by the statutory benefit changes. Among those actually affected, an increase in benefit levels from 80 to 100 percent led to an increase in use of annual sick leave days by about 2. We calculate the implied arc elasticity of sick days with respect to benefit levels to be about 0.9. This is almost identical to what Meyer (1990) finds as an unemployment duration-benefit elasticity. One important implication of our contribution is that calculated elasticities may differ largely, depending on whether they are derived from ITT or ATT estimates.

In addition to elasticities, other highly relevant research questions with respect to the sickness duration-benefit relationship are whether (i) employees mainly adjust short- or long-term absence spells and (ii) whether cuts and increases in benefits trigger the same behavioral responses or not. The mean estimates from this paper are comparable to both studies discussed above, *ZK2010* and *PS2010*. However, while *PS2010* find that both cuts and increases in benefits mainly affect long-term absence spells, our findings differ and suggest that different types of employees responded to cuts and increases in benefit levels. In particular, *ZK2010* show that the lower part of the sickness day distribution responded to benefits cuts. In contrast, this paper shows that mainly higher quantiles of sick leave days responded to the benefits increases, suggesting that employees with longer spells simply extended existing sick leave periods. This difference in the findings between *PS2010* and *ZK2010* is puzzling, but is probably attributable to different research designs. Clearly, more research on the a(symmetry) of behavioral reactions is needed.

The second main research strand in the field of SI asks whether shirking or presenteeism dominate in behavioral reactions. Even though it is very hard to provide reliable evidence in any direction, the answer to this question is critical to evaluate the effects of such public policy programs. In general it is desirable to distinguish intended from unintended use, but, in the case of SI, there are also possible negative external effects that need to be taken into account: If benefits are too restrictive, employees may go to work even though they are sick (*presenteeism*) and spread contagious diseases to co-workers and customers. On the other hand, if benefits are too generous, employees may call in sick although they are healthy (*absenteeism/shirking*), which may negatively affect co-workers and customers as well through various channels. Hence, an efficient SI system minimizes both phenomena.

The third main research question is closely linked to the presenteeism-shirking question and asks about the health effects of SI programs. This paper makes substantial contributions to the second and third SI research area. To begin with, we face a fundamental identification problem when it comes to how to empirically identify presenteeism and shirking. The reason is that the question of whether someone is healthy enough to work is—at least partly—highly subjective and only imperfectly observable. Typically, economists refer to any behavioral change that is triggered by a change in insurance coverage as moral hazard. We try to approach the presenteeism-shirking identification issue in the following way: we interpret it as evidence of shirking when the null hypothesis of no significant health changes cannot be rejected using a battery of different tests. Indeed, and in line with *PS2010*, we do not find any empirical evidence for health improvements, either among pre-reform healthy or unhealthy employees. This finding clearly suggests that presenteeism was not a major problem prior to the reform. On the other hand, the test proposed is pretty demanding, which is why we also complement it with extensive tests of effect heterogeneity. The previous literature has been relatively silent on what type of workers change their behavior. By interacting the reform variable with a rich set of survey covariates, we identify individuals who are particularly sensitive to changes in incentives. The characteristics of these individuals may in turn be indicative of whether sick people or shirkers are generally more sensitive to incentives. The suggestive evidence obtained by this exercise yields mixed results. On the one hand, mainly confident male workers in a partnership extended their sick leave periods, which is suggestive of the shirking explanation. On the other hand, we find that predominately unhealthy employees with many sick days extended their sick leave, which is in line with the presenteeism explanation.

To our knowledge, the fourth and final research area has not been studied at all in the context of SI; namely how employers and the labor market adjust to sick pay employer mandates. We calculate that, as a result of the reform, sick leave payments increased by about €1.8 billion or 8 percent per year. Economic theory suggests that employers would react to such an exogenous shock to labor costs. However, the German labor market is highly regulated, and dismissal protection legislation is among the strictest in the world. Indeed, focusing on the extensive labor supply margin, we do not find evidence that more private sector employees became unemployed in post-reform years, relative to pre-reform years and unaffected sectors of the economy. However, there is evidence that the transition from unemployment to employment—i.e., the hiring process—was negatively affected. Lastly, focusing on the intensive margin, our findings suggest that overtime increased in the affected private sector relative to the unaffected sectors of the economy. Thus, this paper also contributes to the literature on employer-mandated benefits and compensating differentials.

The next section gives a brief literature overview. Section 3 explains Germany's sickness insurance system. Section 4 gives more details about the data. This is followed by a discussion of the empirical estimation strategies in Section 5. In Section 6, we first estimate the causal reform effect on sickness absence behavior. Then, we provide evidence on effect heterogeneity and the underlying mechanisms. The last part of Section 6 calculates the reform effect on sick leave payments and provides evidence on how employers might have reacted to the increase in labor costs. Section 7 concludes.

2 Brief Literature Overview

First, although the literature on workplace absences is quite rich in general, few studies have convincingly identified causal relationships in the context of SI. Using data from a large Italian company, Ichino and Maggi (2000) show that to a large extent cultural backgrounds determine temporary absence behavior. Ichino and Riphahn (2005) find that probation periods have a substantial impact on workplace absenteeism. Using the same data set as this study, Riphahn (2004) shows that the latter finding also holds for Germany. While many other studies analyze the determinants of absence behavior in other contexts (Barmby et al., 1994; Winkelmann, 1999; Barmby et al., 2001, 2002; Thoursie, 2011; Frölich et al., 2004; Engellandt and Riphahn, 2005; Ichino and Moretti, 2009; Barmby and Larguem, 2009; Hassink and Koning, 2009; Nordberg and Røed, 2009; De Paola, 2010; Pouliakas and Theodoropoulos, 2011; Böheim and Leoni, 2011; Markussen et al., 2011; van den Berg et al., 2012; Goerke and Pannenberg, 2012), there is a paucity of empirical findings showing how the *design* of *SI* relates to absence behavior and the labor market. The literature contains very few studies providing evidence on the causal relationship (Johansson and Palme, 1996, 2002, 2005; Henrekson and Persson, 2004; Puhani and Sonderhof, 2010; Fevang et al., 2011; D'Amuri, 2011; Pettersson-Lidbom and Skogman Thoursie, 2012). All of the studies cited above find that employees adapt their short-term sick leave behavior to economic incentives.

Second, in addition to the SI literature, the related literature on WCI and DI provides some guidance. Two studies from the US analyze the impact of changes in benefit levels for WCI: Meyer et al. (1995) find that a 1987 increase in benefit levels led to longer duration of leave, while Curington (1994) find mixed results based on data from the 1960s and 1970s. Issues surrounding DI have also attracted a great deal of attention in the literature (Burkhauser and Daly, 2012, cf.). 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; Kostl and Mogstad, 2012), although there is also evidence that this is not always the case (Campolieti, 2004). Researchers have also studied the DI application process (Burkhauser et al., 2004, cf.) and the decision to apply for benefits within a lifecycle context (Chandra and Samwick, 2006). But compared to WCI and DI, SI 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.

Third, when employers are mandated to provide SI benefits directly or indirectly, the SI literature can be seen as part of the literature on employer mandates. The effects of employer-mandated benefits on labor market outcomes are analyzed in seminal contributions by Summers (1989) and Gruber (1994). These mandated benefits may be an efficient way to address market failures, provided workers value the benefits highly enough to be willing to accept a reduction in wages. Thus, if we observe a reduction in wages after an employer mandate has been introduced, it indicates that the deadweight loss of the policy is limited. On the other hand, one aim of the policy analyzed in this paper was precisely to redistribute resources, and, from this point of view, a reduction in wages would be an indication that the policy has failed.

The empirical evidence concerning employer-mandated benefits is mixed. Several papers analyzing maternity leave regulations fail to confirm theoretical predictions (Waldfogel, 1999; Hashimoto et al., 2004; Baum, 2006). Studies from Germany do, however, tend to find long-term effects on women's wages (Schönberg and Ludsteck, 2007). The evidence concerning employer-mandated health insurance is equally mixed. Using a natural experiment from Hawaii, Buchmueller et al. (2011) find no evidence that a mandate reduces wages and employment probabilities. On the other hand, Kolstad and Kowalski (2012) find substantial wage shifts in response to the Massachusetts health reform.

3 The German Sickness Insurance System and Policy Reform

3.1 The Sick Pay Scheme and Monitoring System

Before the implementation of the 1999 reform, every German private-sector employer was legally obligated to pay 80 percent of foregone wages for up to six weeks per sickness spell.[2] Obviously, self-employed people are not eligible for employer-provided sick pay. Public sector employees were 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 consecutive weeks, the doctor must issue a certificate

[2] 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.

of long-term illness. From the seventh week onwards, sick pay is provided by the employee's health insurance company 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.[3] In 2011, about 2,000 full-time equivalent and independent doctors worked for the medical service and examined 1.6 million cases of absenteeism (Medizinischer Dienst der Krankenversicherung (MDK), 2012).

3.2 The Policy Reform

In the 1998 election campaign, both the Social Democrats and the Green Party promised to increase federally mandated 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 in long-term sick pay, short-term sick pay was decreased from 100 to 80 percent of forgone gross wages. Ziebarth and Karlsson (2010) (*ZK2010*) analyze the effect of the cut in short-term sick pay and find that it increased the share of employees with zero absence days by about 8 percent.[4] The majority of Germans perceived the cut in sick pay as unfair and socially unjust resulting in a number of strikes 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.[5]

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. 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 declared that 13 out of 27 million employees would receive 100 percent sick pay (Jahn, 1998).[6] In 1997, a poll among craftsmen's businesses showed that 51 percent were voluntarily

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

[4] 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. Moreover, they do not compare the robustness of the results with respect to various DID estimation approaches. The estimated change in sick leave behavior is, however, in line with the findings of the present study.

[5] 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*.

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

providing 100 percent sick pay, probably due to the close relationship and mutual trust between employers and employees in these small companies (Ridinger, 1997). Like all of the other studies that have evaluated the impact of changes in federally mandated 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. We will discuss our definitions of different types of treatment effects in more detail below.

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).

4 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.[7] 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.

4.1 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. However, there is no reason to expect that measurement errors should systematically

[7] 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.

differ between our treatment and our control group. 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.[8] 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*, 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.

4.2 Treatment and Control Group

In our baseline specification, the treatment group consists of all private-sector employees. The control group incorporates public-sector employees and the self-employed—all those who did definitely not experience a legal change in their minimum 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 15,114 observations in the treatment group and 7,877 observations in the control group. A difference-in-differences (DID) analysis based on these two groups gives an estimate of the intention-to-treat effect (ITT); i.e., the effect of raising the legal minimum replacement rate for private sector employees in Germany. This parameter is of obvious interest from a policy perspective, since it captures the effects of the type of changes that the government can actually carry out. Besides, identifying ITT effects typically requires fewer assumptions than any other type of treatment effect.

In a second step, we take imperfect compliance into account. In this part, we focus on industries for which the degree of compliance is known. After an extensive review of

[8] 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 that 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.

all collective agreements in the main industries, we were able to identify some industries (e.g. construction and agriculture) where the pre-reform minimum of 80 percent applied to all workers. Likewise, public sector workers obviously had a 100 percent replacement rate throughout. Moreover, we identified seven main industries where the replacement rate depended on the collective agreement coverage pre-reform: in, e.g., the chemical industry and credit and insurance industries, workers covered by collective agreements had no changes in their replacement rate, whereas all other workers experienced an increase from 80 to 100 percent. Finally, there were some industries for which the rules applying before 1999 could not be recovered—we drop these industries from the "full compliance analysis."[9]

In the full compliance analysis, the treatment variable is a fraction that represents the proportion of workers in an industry receiving 100 percent sick pay. This fraction varies between industries depending on a) the number of workers covered by collective bargaining agreements and b) whether a 100 percent replacement rate was negotiated in these agreements in the aftermath of the previous reform. For most workers and years, this variable will be either a zero or a one—i.e., we know exactly the replacement rate that applies—but in some few cases it will be a fraction since the replacement rate depends on the collective agreement status of the worker. Since we do not have information on the collective agreement status of individual workers, there will clearly be some degree of measurement error prior to the reform within the group where different rules applied to different workers.

The analysis in this second part delivers an estimate of the average treamtent effect on the treated (ATT). This parameter is also of great interest, as it shows the individual-level labour supply adjustment. Thus, our analysis delivers two sets of parameters—the first one related to the policy question of the overall response to the reform, and the second one to the individual labor supply reaction. Moreover, this dual approach allows us to gauge the validity of our approach, since we may check whether the two sets of estimates are consistent with each other.

5 Estimation Strategy and Identification

5.1 Identification of Causal Effects

As will be discussed in more detail below, we utilize a rich set of socioeconomic background information to make the treatment and the control group in the ITT analysis—i.e., private sector employees vs. public sector employees and self-employed—as comparable as possible. We apply matching methods and obtain two samples that are almost identical in terms of observables.

However, the crucial identifying assumption in any 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, for the following reasons, we believe that it is very likely

[9] Puhani and Sonderhof (2010) (*PS2010*) evaluate the same reform but select the sample in a very different way. Their identification strategy differs substantially from ours. For example, they do not differentiate between the intention-to-treat and the full compliance effect. Most importantly, *PS2010* define the control group as those public *or* private sector employees who claimed in 1995 that their wage would be determined by collective bargaining and who remained in the sample until 2000. Hence they assume that employees who underlay a collective agreement in 1995 were not treated and everybody else was treated. Moreover, for the pre-reform years, they require that respondents had not changed jobs since 1995, but they do not make this requirement for post-reform years.

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 employee cost of workplace absence, an outcome that we are able to observe directly. Since the reform was a reflex 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 campaign promise (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. In a robustness check, we weight the sample with the individual inverse probability of not dropping out in post-reform years. In addition, it might be the case 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 the case 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. Moreover, in the final part of the paper, we investigate transition from unemployment to employment in the different sectors and vice versa.

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 1, 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 3, 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.[10] 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.

[Insert Figure 1 about here]

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. The most plausible explanation for the closing of the gap from 2001 to 2003 is a recession at the beginning of 2002. It is a well-documented stylized fact that private sector absenteeism rates and unemployment rates are negatively correlated

[10] Please note that Figure 1 is not comparable to Figure 4 in Ziebarth and Karlsson (2010). As discussed in more detail in the Conclusion, this is because the sample selection and the definition of the treatment and control groups differ. More precisely, in Ziebarth and Karlsson (2010) the small and homogeneous sample of respondents with long-term sick leave spells are excluded (see p. 1111 in Ziebarth and Karlsson (2010)). Also, remember that the average sick leave rate of the control group is a weighted average of the average sick leave rate of public sector employees and self-employed. It is a coincidence that the weighted average of these two groups is about the same as the average sick leave rate of private sector employees in pre-reform years under the 80 percent replacement regime.

(Askildsen et al., 2005). Note that the absenteeism gap closed due to a *decrease* in the private sector absenteeism rate between 2001 and 2003. In this time period, the German labor market saw a sharp 1.1 percent increase in unemployment rates as illustrated by the unemployment rate graph in Figure 1. While, between 1997 and 2000, there were relatively stable economic conditions with a smooth decline in the unemployment rate, we observe an abrupt 1.1 percent (or 500,000 people) jump in unemployment from 2001 to 2003. However, it should be kept in mind that Figure 1 paints a raw, unconditional picture. In the empirical assessment below, we focus on the years 1997 to 2000 and correct the sample composition with respect to a rich set of covariates.

In recent years, the drawbacks and limitations of DID estimation are extensively debated. 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 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 ($16\times4 = 64$ 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).[11]

While the econometric literature on standard errors is large and influential, a more recent paper emphasizes another crucial and potential source of bias in the context of DID estimation: Hong (2012) shows that compositional changes in the treatment and control group over time might have a substantial effect on the estimates. He uses the example of Napster and its impact on music sales to illustrate this point. Since the treatment is internet technology, the composition of the treatment group—those with access to broadband internet—substantially changes over time. Hong (2012) develops an identifying restrictions test for a DID estimator under compositional changes. This is an extremely useful test, especially when researchers are restricted to the use of cross-sectional data. Fortunately, we can exclude that compositional changes bias our estimates. First, in our setting, a whole sector of the economy represents the treatment group. Compositional changes are less of an issue here. According to a balanced version of our data, only 2.1 percent (172 people) switched from the private to any other sector in pre-reform years. In post-reform years, the figure remained very stable (1.6 percent). Looking at the other direction—at transitions from public sector and self-employment to the private sector—the transition rate also almost exactly remained the same in pre- and post-reform years (2.39 vs. 2.37 percent). Second, since we rely on panel data, we can formally test whether compositional changes affect our results. As mentioned above, we implement robustness checks that drop job, industry, and sector-changers.

As discussed in Section 3.2, even before the increase in statutory sick pay, some employers agreed in collective bargaining to voluntarily provide 100 percent sick pay. We cannot precisely identify employees who were subject to such collective wage agreements. Our approach is to combine two different strategies. In a first approach, we focus on intention-to-treat effects (ITT), which is mostly applied by researchers when analyzing changes in statutory minimum standards: the *overall* effects of changes in statutory sick pay are evaluated. In contrast to other countries, 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 a second approach, we focus on industries for which we know the degree of implementation of the reform. We are still unable to identify pre-

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

reform exposure to treatment at the individual level in some industries, but for most, we at least know the fraction of affected individuals. 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.

5.2 Estimation Methods for Assessing the Causal Reform Effects on Workplace Absences

Parametric DID Approaches

OLS

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

$$Y_{it} = \lambda T_t + \pi D_{it} + \theta D_{it} \times T_t + X_{it}\psi + \rho_t + \phi_s + \epsilon_{it} \tag{1}$$

where Y_{it} stands for the annual number of days of sick leave for individual i in year t, T_t is a post-reform dummy, D_{it} is the treatment group dummy, and $D_{it} \times T_t$ 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 provided the assumptions required in the DID approach are satisfied. 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 $1 \times K$ column vector X_{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 4.1. 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 (cf Delgado and Kniesner, 1997; Cameron and Trivedi, 2005; Winkelmann, 2008). Thus, we use this model in all count data specifications.

Matching DID Approaches

In addition to these parametric estimates, we also consider a set of different matching versions of the DID estimator. The main reason for using non-parametric or matching techniques is the concern that the functional form assumptions underlying the previous specifications may be incorrect. This is true by default for the linear specification, and it may also be the case for our ZINB specification. On the other hand, the matching approaches have their own issues. One such is whether to take first differences of the outcome and then match, or to match in two dimensions. In our case, since the panel used so far is four years long and unbalanced, it seemed natural to first match and then calculate differences.

Thus, the estimand is now the average conditional treatment effect on the treated (ATT):

$$\begin{aligned} ATT(x) = \mathbb{E}\left[Y_{it} \mid D=1, T=1, X=x\right] - \mathbb{E}\left[Y_{it} \mid D=1, T=0, X=x\right] \\ - \mathbb{E}\left[Y_{it} \mid D=0, T=1, X=x\right] + \mathbb{E}\left[Y_{it} \mid D=0, T=0, X=x\right] \end{aligned} \tag{2}$$

where expectations have been taken with respect to the distribution of covariates in the treatment group. In a first step, we estimate the propensity score (PS) for $D = 1$ by means of a logit model and select a subset of X as covariates for the propensity score (Part A of Table 1) using likelihood ratio tests on zero coefficients. First, we conduct the test for control variables in levels and in a second step for their interactions (Imbens, 2008). One issue highlighted by Hong (2012) is that it is not obvious whether the PS should be estimated based on pre-intervention observations, based on the entire sample, or whether one should use a combination of pre-intervention and post-intervention propensity scores. The choice of approach can matter a lot whenever there are compositional changes within the groups over time. In our case, the problem would of course be easily solved by balancing the panel. However, since we decided to use an unbalanced panel in this analysis, the potential problem of compositional changes naturally needs to be considered. Hong (2012) also presents a useful test of the assumption that there are no compositional changes that may threaten the identification strategy; we implement this test (see Section 6.1 for more details).

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 seen in Figure 2. The PS distribution for both groups shows a large overlap with the region of common support lying between PS values of 0.12 and 0.96.

[**Insert Figure 2 about here**]

The first matching method that we employ is stratification matching or blocking. Based on the estimated PS for $D = 1$, the sample is cut into blocks such that the covariates are balanced within each block. Then, block-by-block average treatment effects on the treated are obtained by calculating the sample equivalent of the treatment effect $ATT(x)$. 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) shows 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.

A second matching method is k-to-one nearest neighbors matching with replacement. In our case, each observation with DID=D×T=1 is matched with five observations from each one of the three other groups. Next, the double difference in the outcome variable is calculated (Heckman et al., 1998; Lechner, 2002).

Matching on the Treatment Group Indicator D to Improve Balancing Properties

Both regression and matching methods have drawbacks. If treatment and control group differ substantially in their observed characteristics, then parametric approaches use the covariate distribution of the comparison group to make out-of-sample predictions. Imbens and Wooldridge (2009) propose to evaluate differences in covariates for treatment and control group by the scale-free **normalized difference**:

$$\Delta_X = \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{\sigma_1^2 + \sigma_0^2}} \tag{3}$$

with $\bar{X}_1$ and $\bar{X}_0$ denoting average covariate values for the treatment and comparison group, respectively, and σ standing for the sample variance of X. As in our case there are three distinct comparison groups (see equation 2), it is not obvious which the relevant statistic for Δ_X is. However, since covariates change only slowly over time, we decided to make the comparison between the treatment group (including data points before and after the intervention) and observations belonging to the control group not exposed to treatment. As a rule of thumb, a normalized difference exceeding 0.25 is likely to lead to sensitive results Imbens and Wooldridge (2009).

Thus, applied to our case, we 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. Considering the "raw sample", we see that there are a couple of variable that are close to being problematic—in particular *immigrant* status and *13 years of schooling*. In general, however, the two groups are very similar in their observable characteristics.

[Insert Table 2 about here]

We now apply two different matching procedures to improve the balancing properties across the treatment and control groups. Using combined matching and regression approaches (see next subsection) requires this as a first step. In the second step, one applies regression approaches to these matched samples. Note that the first step—balancing covariate distributions—requires that we match on the treatment group indicator D. Columns (3) to (6) show the "blocked sample" and the covariates' mean values for the treatment and control group plus the normalized difference for observations within the two groups that have a common support. Blocking improves the balance of the covariates between the treatment and the control group. In columns (7) to (9), we present the corresponding statistics for a "matched sample", which we obtain using five-to-one nearest neighbors matching. The matched sample shows better balancing properties than the raw sample and the majority of the normalized differences are now well below 0.10.

Combining Matching and Regression DID Approaches

Even for the matched and the blocked samples, small differences between treatment and control group remain. These differences may lead to biased estimates. Abadie and Imbens (2011) show that the simple nearest neighbor matching estimator includes a bias term, which leads to inconsistencies and should be corrected for. Thus, Imbens and Wooldridge (2009) propose two approaches that both combine the strengths of parametric and matching estimators.

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 (3) to (6) of Table 2. In the second step, parametric regressions—as detailed in Section 5.1—are run within each block. Then, the within-block treatment effects are weighted by the number of treated individuals in each block 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 (2011). 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. Then, in the second step, the counterfactual potential outcomes are calculated, based on regression-adjusted values from the three groups not receiving treatment (Abadie et al., 2004):

$$\hat{Y}_i^{td} = \frac{1}{M} \sum_{j \in \Gamma_M(i)} Y_j + \beta_{td}' \left(X_i - X_j \right) \ \forall t, d : t \cdot d = 0 \tag{4}$$

where $\Gamma_M(i)$ denotes the matches for unit i.

As for the variance of the estimated treatment effect, Abadie and Imbens (2011) suggest an estimator that is based on nearest-neighbor matching within the own treatment group. In a recent contribution, Hanson and Sunderam (2012) extend this estimator so as to additionally correct for unobserved common group errors. We implement the clustered variance estimator proposed by Hanson and Sunderam (2012). However, it turns out to make little difference in practice.

6 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. We first show how increasing the generosity of the sickness insurance system has causally affected sick leave behavior and compare intention-to-treat (ITT) with full compliance estimates. Then, we provide evidence on the underlying mechanisms by looking at effect heterogeneity and the reform's impact on employee health and well-being. Finally, we calculate how the expansion of federally mandated benefits affected actual employer sick leave benefit payments and provide evidence on labor market adjustments to the exogenous shock in labor costs.

6.1 Assessing the Causal Reform Effects on Sickness Absence

Parametric DID Approaches

We start by estimating parametric OLS-DID and ZINB-2-DID ITT models using the raw sample of Table 2 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 1.38 that is statistically significant at the 5.3 percent level. The ZINB-2 model gives an estimate of 1.10 with a standard error of 0.47. The unconditional double difference of the means of the two groups for the two time periods is 1.49 (std. err. 0.73; not shown) and very close to the OLS estimate in column (1), which reinforces the credibility of the common time trend assumption.[12]

[Insert Table 3 about here]

[12] This finding demonstrates that correcting for observables does not affect the estimates. We consider this as suggestive evidence that unobservables are unlikely to significantly confound our estimates either.

Matching DID Approaches

Columns (3) and (4) give the results when two different matching variants are applied using one of the matched subsamples of Table 2. In column (3), the blocked sample— based on the propensity score (PS) for $D = 1$—is used. By comparing the average values of the four groups as in equation 2 within each block, the block-specific reform effects are calculated. Finally, they are aggregated to a weighted overall average. This method produces an estimate of 2.28 with a standard error of 0.76.

Column (4) yields the estimate when five-to-one nearest neighbors matching is applied using the matched sample of Table 2. The estimated reform effect is 2.02 and significant at the five percent level.

In column (5), we present results from a bivariate local linear matching procedure—a procedure where the local linear regression is weighted according to a bivariate kernel function based on pre- and post-reform propensity scores.[13] In this specification, the estimated effect is 2.24 and it is significant at the one percent level.

Combining Matching and Regression DID Approaches

According to Imbens and Wooldridge (2009), the most suitable methods combine regression and matching and, consequently, are more flexible and robust than other methods. Column (6) 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 1.49 and significantly different from zero, is obtained as an average of the within-block estimates weighted by the block size of the treated. 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 5.1. The resulting estimate is similar to the one in column (6) and yields a reform effect of 1.63 (std. err. 0.33).

We conclude that the estimates do not differ very much in magnitude, that all estimates carry the expected sign, and that all estimates are statistically different from zero. The size of the coefficients varies between 1.1 and 2.2 and almost all confidence intervals overlap. These findings suggest that the identified effect is robust and not very sensitive to the functional form imposed. All in all, the conventional and transparent OLS-DID model does a relatively good job of estimating the effect of the reform. Thus, in the following, we focus on conventional OLS-DID models.

If we take the mean number of absence days in the pre-reform period for the treatment group, which was about 10, and relate the lower bound ZINB-DID estimate of about 1 additional absence day per employee and year 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. A reform effect of 1.5 days would yield an increase of 15 percent. Note that these are ITT reform estimates. As discussed previously, among the reform's target group, only about half of all employees effectively experienced an increase in sick pay. Hence our ITT estimates suggest that the employees actually affected increased their days of sick

[13] The reason for considering this specification was that we implemented the test provided by Hong (2012). It is equivalent to the "DDM estimator" used in Hong's paper. The null hypothesis of no compositional changes over time was rejected. However, the results using this bivariate matching procedure are no different from those based on univariate matching. Our interpretation is that it is not primarily compositional changes within groups that give rise to the rejection of the null—but rather discrepancies in covariate distributions between the treamtent and the control group. Thus, the doubly robust models provide more convincing evidence than these simple matching results.

leave by about two days per year. Indeed, this assumption is almost exactly verified in the "full compliance estimation approaches" presented in the next 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. We identified industries that fully complied with the federally mandated minimum standards by providing only 80 percent sick pay in the pre-reform years. We also identified industries 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 Section 4.2 as well as the notes to Table 4.

[Insert Table 4 about here]

The last four 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 four columns. Columns (1), (2) and (5), (6) make use of the standard intention-to-treat (ITT) 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 the ITT models was effectively affected by the reform, we expect the estimated behavioral effects to be smaller than the ones in the full compliance analysis. The even-numbered columns display Fixed-Effects (FE)-DID models, while the odd-numbered columns show OLS-DID models.[14]

The main findings from Table 4 can be summarized as follows: First, the estimated OLS "full compliance effect" of the expansion is around 2 days, which corresponds to an increase in absence days of about 20 percent (columns (3) and (7)). As expected, it is larger than the effect of the ITT-approach (columns (1) and (5)). The implied arc elasticity of these models 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.

Second, the FE point estimates are systematically about twice as large as the OLS point estimates. One explanation could be that individual time-invariant unobservables downward bias the OLS estimates. A second explanation would refer to how the effects are identified. While the OLS-DID models exploit the full range of the unbalanced sample, the FE-DID models rely on individuals who are observed at least twice and have a change on their DID=D×T variable. This means that the FE-DID model is mainly identified by full-time employees with a stable employment history and who do not drop out of the sample. In contrast, the OLS-DID model also includes employees with less stable employment relationships, employees on temporary employment, or on probation—all of whom are more threatened by unemployment and have been shown to react less to sick leave incentives (cf. Ichino and Riphahn, 2005).

Third, the results between the full sample and the private sector sample are very robust. Please note that—in addition to Table 3—Table 4 provides us with two new distinct samples, alternative treatment and control groups, estimates ITT and full compliance effects, and

[14] Note that we always report the full sample size, although the Fixed-Effects results are only identified by individuals who change their DID=D× T variable status over time.

compares OLS-DID and FE-DID models. Considering this, the labor supply estimates are indeed 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 ITT approach. Apart from having analyzed the sensitivity of the ITT results with respect to various regression, matching as well as combined DID methods, more robustness checks are shown in Table 5.

To further test the differences between OLS-DID and FE-DID models, column (1) of Panel A estimates the same model as in column (1) of Table 3, but uses fixed-effects regression techniques. Again, exactly as in Table 4, the highly significant point estimate doubles in size to about 2 (days). The consistency of this finding is reassuring. Remember that we obtain this pattern for three different samples with varying treatment and control groups. As mentioned above, it should be kept in mind that the FE-DID model is solely identified by respondents who are observed working full-time at least twice and who changed their treatment status as indicated by the difference-in-differences indicator DID=D×T. This excludes employees with less stable employment histories—those who have been show to be, *ceteris paribus*, less responsive to monetary sick leave incentives (cf. Ichino and Riphahn, 2005). On the other hand, we should also keep in mind that the OLS-DID and FE-DID estimates still lie within the same confidence intervals.

Column (2) of Panel A checks whether panel or labor market attrition might drive our results. We weight the estimates with the inverse individual probability of not dropping out of the sample in post-reform years. The size of the statistically significant estimate increases to 1.9. This finding supports the hypothesis that we find larger FE than OLS point estimates because the effects are identified by different individuals in our sample.[15]

In column (3), we compare only 1997 with 2000 to exclude the possibility that anticipation or adaptation effects might drive our results. Since the sample size reduces substantially, we lose power and the estimate is only significant at the 10.9 per cent level. However, it is of similar size as the baseline estimate in column (1) of Table 3.

Column (4) of Panel A 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 1.6.

The last column of Panel A 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.

The first three columns of Panel B deal with concerns that treatment-related compositional changes and selection into occupations might drive or bias the results. Column (1) excludes all individuals who changed their employer at least once and who belonged to the treatment group for at least one year. The estimate is significant at the 10 percent level and almost identical to the standard estimate in magnitude. Column (2) excludes all private-sector employees who changed the industry in the post-reform period. Given the reform design, it is likely that collective bargaining assured that sick leave regulations only varied across, but not within, industries. Again, the size of the estimate is close to the main estimate

[15] Again, please note that although the sample size reported for the FE-DID model formally equals those of the OLD-DID model, the FE-DID model is only identified by individuals who are observed at least twice and change their $DID{=}D{\times}T$ status over time. For example, SOEP Sample F contains 11,000 respondents. This sample was drawn in the year 2000—with absence information about 1999. The OLS-DID model includes respondents from this refreshment sample, while the FE-DID model does not.

in the first column of Table 3 and is marginally significant at the ten percent level. Column (3) excludes sector changers, i.e., employees who switched between the treatment and control group, and additionally weights the regression with the inverse probability of being observed in post-reform years. Consistent with the findings above, the estimate is highly significant at the 1.7 percent level and about 2 days.

In another check (results not displayed in Table 5), we look at whether the reform was followed by a change in the rate of job switches. 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 is almost identical 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.[16]

As a final check (results not displayed in Table 5), we look at whether the rate of changing sectors—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 and 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 (4) and (5) of Panel B 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.

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]

6.2 Assessing Effect Heterogeneity and Health Effects

In the following two subsections, we shed more light on the underlying mechanisms that drive the average labor supply effects found above. For this purpose, we make use of rich socio-economic background information and provide evidence on heterogeneity in the reform effects. Then, we provide empirical evidence on whether the increase in sick leave improved employee health or well-being.

Heterogeneity in Effects: Who Reacted to the Increase in Generosity?

Table 7 displays extensive tests on treatment effect heterogeneity. Every column shows one OLS-DID model as in the main specification in column (1) of Table 3. The only difference is

[16] All results that are cited here, but are not displayed in Table 5, are available upon request from the authors.

that the corresponding variable—with which we want to perform the heterogeneity test—is included both in levels *and* as a triple interaction interacted with the DID=D×T 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 already included in the standard model and vector X_{it}, we interact the dummy variable *female* with D×T and run the model under the inclusion of this additional triple interaction term.[17] The triple interaction term tells us how females reacted to the reform, relative to men. In this case, the $D\times T\times female$ point estimate is imprecisely estimated, it is negative, and relatively large in magnitude (-0.85). This provides evidence that women did not react as strongly to the increase in sick leave benefits as men did. This finding is reinforced when we apply an alternative approach and run our standard OLD-DID model explicitly on the female and male subsample (results not shown in Table 7). In the former case, we obtain an insignificant 0.66 reform effect, and in the latter case, a significant reform effect of 1.55 days. The finding that the increase in workplace absences is mainly driven by men is in line with Ichino and Moretti (2009), who suggest that men are more prone to shirking behavior than women.

Panel A of Table 7 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. Both triple interaction terms are not only insignificant, but also very small in magnitude. There is evidence that the richer half of the population reacted less than the poorer half—the -1.4 triple interaction term is significant at the 7 percent level. From column (4), we infer that 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. This is in line with findings from Goux et al. (2012), who study the statutory reduction in weekly working hours in France and provide evidence on spousal leisure complementarity.

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.[18] Precisely how the six dummy variables are generated is explained in the notes to Table 7. The findings concerning employees in good health are mixed. Column (1) shows that the 12 percent of the sample in the highest SAH category reacted as the rest: the triple interaction term is close to zero in size. Column (3) stratifies on those 10 percent in the highest health *satisfaction* category—a more subjective and slightly different health measure. The triple interaction term is relatively large (-1.3) and significant at the 5.5 percent level. If we restrict the sample to solely those in the highest

[17] Please note that we always control for our standard set of covariates simultaneously, as indicated in equation 1. An even more rigorous way to conduct the heterogeneity analysis would be to add the further interactions, e.g., $T \times female$ and $D \times female$ in column (1). However, these additional interaction terms turned out insignificant in almost all cases. In further tests whether the parameters associated with these additional terms were jointly zero, we were unable to reject the null in all but a couple of cases. Another possibility would be to include *all* triple $D \times T \times [covariate]$ interaction terms simultaneously. However, due to substantially varying sample sizes in panels C and D this is not feasible. In case of Panel B, it would be unclear how to interpret a $D \times T \times [healthverygood]$ triple interaction term, when simultaneously controlling for 5 other health-related triple interactions—in addition to controlling for the 6 health covariates in levels. However, to be on the safe side, in Panel A and B, where we make use of our standard set of covariates, we always include a whole set of six triple interaction terms that we obtain by interacting D×T× [*the covariates in the six column headers of Panel A*]. The results are almost identical to running separate regressions with just one triple interaction term.

[18] Note that, in Germany, certified disability does not necessarily imply eligibility for DI benefits. About 4.4 percent in our full-time employed working sample are disabled (see Table 1).

health satisfaction category (2,150 obs., results not shown), we find a precisely estimated, but insignificant and small 0.14 reform effect. Hence, it is safe to conclude that this subgroup did *not* take more days off due to the reform. The last measure of good health—the 78 percent of the sample who claimed that they were not impaired by their health status in their everyday life—yield an imprecise result (column (5)). However, if we explicitly select on this sample (17,999 obs., results not shown), the reform effect is 0.9 days and significant.

The findings concerning employees in a bad health shape are less ambiguous—stratifying on those 8 percent in the worst two SAH health categories yields a precise and large estimate of about 5 additional sick leave days. The triple interaction term for the second bad health measure (low health satisfaction, column (4)) is even larger and significant at the 12 percent level. Explicitly selecting on these 5 percent of the sample (1,214 obs., results not shown) and running our standard OLS-DID model yields a reform effect of 13 days which is significant at the 6 percent level. This finding is reinforced by the last column, where we find that those who are officially certified as disabled took 9.5 more days off and drove a good deal of the average behavioral reaction. As a final check, including the lagged number of sick leave days in levels and as triple interaction reveals that the level effect is 0.3 and significant, but there is no evidence the the reform itself triggered more sick leave days in t when employees had sick leave days in *t-1* (not shown).

We also run DID Quantile Regressions and estimates suggest that employees with an above-average annual number of sick leave days took more days off as a reaction to the reform: we find significant behavioral reactions in the 85^{th} and 90^{th} percentile of the annual sick day distribution – assuming no reranking of workers, this means that employees with between 15 and 22 annual sick days called in sick 1 to 2 additional days as a result of the reform. For the sake of saving space, we do not report the detailed results which are available upon request.

[Insert Table 7 about here]

In Panel C we exploit objective and subjective workplace characteristics. Column (1) shows that people in small firms with fewer than 20 employees called in sick more often as a result of the reform. Column (2) finds the same for workplaces without a work council. These findings support and reinforce our identification strategy, since employees working for such firms were exactly those who were most likely to actually experience increases in sick pay, given the reform implementation process (see Section 3.2).

The triple interaction term for blue collar workers is small in size, which suggests that blue collar workers did not react differently from white collar workers (column (3)). There is some (imprecisely estimated) suggestive evidence that workers in downsizing companies did not react to the reform, probably due to fear of unemployment. However, interestingly enough, the 4 percent of employees who more or less knew in 1999 that they would lose their job in the next 2 years (subj. prob. > 80%, see notes to Table 7), increased their absence by 8.5 days more than the rest. Note that the large behavioral reaction could either be a consequence of the job loss or the cause, since the question was only asked in 1999. However, laying off a worker because they use too many sick days is illegal in Germany.

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, of surprisingly high magnitude, and significant at the 12 percent level. This effect even increases in size and becomes significant at the 1 percent level if we run the model on the 6 percent subsample of employees with such an attitude. That attitudes matter in this unexpected way may be interpreted as evidence for shirking behavior. Remember that we control for a rich set of background information and also for individual health.

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. In our interpretation, extroverted and self-confident employees predominantly agree with these statements. These employees probably take action more easily and do not fear to call their employer to call in sick.

All in all, we find strong evidence of a substantial degree of heterogeneity in the behavioral responses to the increased sick leave generosity. For some stratifying variables, we find behavioral reactions of up to 10 times the mean reaction. 14 out of 24 models either carry significant or very small triple interaction terms. Four models have marginally insignificant triple interaction terms, but carry significant reform effects when we select on the subsample of interest. If we wanted to characterize—in a slightly oversimplifying manner—the typical employee who took more sick leave days, it would be a self-confident male in a partnership. In addition, small subgroups of about 5 percent of the sample size—those who expected to lose their jobs, those with disability certificates, and those with very low health satisfaction—showed extremely large reactions and decreased workplace attendance by about 10 days per year. In general, the upper tail of the distribution—those with between 15 and 22 annual sick days—showed behavioral reactions and called in sick 1 to 2 additional days as a consequence of the increase in sick pay.

Although the results of this heterogeneity exercise do not unambiguously settle the issue of shirking versus presenteeism, we consider them as suggestive evidence. Taking the oversimplification for granted, we would conclude that, among all employees, the share of potential shirkers lies between 10 and 20 percent since self-confident males in a partnership account roughly for that share. The share of very unhealthy employees, who might drive presenteeism, lies between 5 and 15 percent. This is in line with recent self-reported evidence from German surveys that specificly asked respondents about their shirking and presenteeism behavior. According to these surveys, 15 percent of employees took their last day off for other reasons than personal sickness and up to 40 percent claimed that they went to work sick at least twice in 2008 (Böcken et al., 2009; Aon Consulting, 2010).

Health Effects: Did More Sick Leave Improve Employee Health?

If presenteeism was widespread prior to the reform, is it possible that increasing insurance coverage decreased the number of employees who went to work despite being sick? At least it would be a main economic justification for more generous sick leave policies since the spreading of contagious diseases to co-workers and customers induces negative external effects. This argument is often cited by advocates of more generous sick leave policies who also claim that overall workplace productivity would actually increase as a result. Opponents mainly cite shirking behavior and negative employment effects as one potential undesirable consequence.

The finding that employees with many sick days, those with a disability certificate and/or with low subjective health satisfaction reacted very disproportionately is in line with the presenteeism argument. However, if presenteeism significantly decreased as a result of the reform, then one might also expect to find an improvement in employee health. In fact, while the heterogeneity test evidence above is only suggestive for the existence of absenteeism or presenteeim, individual health measures allow us to implement a test that may help to quantify the relative importance of reduced presenteeism in the overall reaction to the reform. This test is based upon the rather strong assumption that a significant change in presenteeism—as a reaction to changes in sick leave benefits—would lead to a significant change in health and thus in our health measures.[19]

In Table 8, we provide evidence on this by running the same OLS-DID models as before, but now using three different measures of poor health as well as subjective well-being as the *outcome* variable. In addition, in panels B and C, we focus on health heterogeneity effects and provide results for those who had either more than 20 or less than 5 annual sick leave days in one of the two pre-reform years. To take account of health-related labor force composition effects, we balance the samples in Panel B and C.[20]

[**Insert Table 8 about here**]

Table 8 illustrates that we do not find any empirical evidence that employees' health improved as a result of the increase in public insurance coverage. All estimates are very close to zero and insignificant. This finding is robust to three different subjective measures of bad health. Moreover, we do not find evidence for health effect heterogeneity and neither significant health changes for employees in pre-reform bad nor in pre-reform good health. Also, there is no evidence that employees' general well-being improved. Since most of the estimates have small standard errors, this test clearly supports the shirking explanation.

Critics could argue that our health measures would not capture a reduction in spread of contagious diseases appropriately. This is, however, unlikely since we (*i*) employ four different health and well-being measures; (*ii*) stratify the sample in different ways; (*iii*) rely on health measures conventional in the economic literature; and (*iv*) would argue that a significant reduction in flu epidemics should manifest itself in these subjective health measures that capture the transitory *and* permanent health of the respondents who are mainly interviewed in the first quarter of a year. For example, *SAH* asks about the "current" health. *Health satisfaction* refers to "How satisfied are you **today** with our health?" and clearly incorporates a transitory health component. In contrast, the *health impairment* measure mainly captures the health stock, as the question reads: "**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?"

6.3 Assessing Labor Market Adjustment Mechanisms

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

[19] In our opinion, the theoretical link between sick leave benefits and health care utilization is less obvious. First, finding an increase in hospitalization rates as a consequence of increases in sick pay would suggest that employees who would have otherwise worked (sick) now hospitalize themselves. We believe that this is highly implausible, especially since, second, it has consistently been shown that the price elasticity for inpatient care is very low and lies around -0.2 (Manning et al., 1987; Ziebarth, 2010). In addition, we need to consider that the average net wage in the treatment group was about €1,600 or €53 per day. The 20 percentage point increase in benefits equals an average increase in benefits of about €11 per day. At the same time, daily copayments for inpatient stays amounted to about €7.50 for almost everyone in the treatment group in 1998 and 1999. Lastly, if any, we might expect an increase in outpatient visits as a consequence of the reform since employees on sick leave must have doctor certificates from the fourth day of their sickness. Empirically, we do not find evidence that health care utilization increased significantly in post-reform years. Exploiting the number of hospital nights in the last calendar year as outcome variable, we obtain a non-significant coefficient estimate of -0.0007 or -0.1 percent. Omitting the year 1998 and exploiting the number of doctor visits in the last quarter as outcome variable, we obtain a non-significant coefficient estimate of +0.0416 or +2.3 percent.

[20] The findings are robust to using alternative measures of bad and good pre-reform employee health. They are also robust to not balancing the subsamples or weighting the regression with the inverse individual probability to not drop out of the sample in post-reform years. All results are available upon request from the authors.

perspective. In the final part, we want to present empirical evidence on how expanding employer benefit mandates might affect firms and induce changes in the organization of and demand for work.

By How Much Did Employer-Provided Sick Leave Payments Increase?

First, we assess how the increased obligation to provide sick leave benefits affected employers' sick leave payments, and thus labor costs, using a simple accounting model. The overall increase in employer-provided sick leave payments within the treatment group 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.[21] This is illustrated in equation (5) below:

$$\Delta SLP = \sum_{i=1}^{N} (Y_{i1}W_{i1} - 0.8 \times Y_{i0}W_{i0}) \tag{5}$$

where SLP stands for Sick Leave Payments, W_{it} is the Daily Gross Wage, and N is the total size of the private sector workforce. As can be seen, for the pre-reform period, we assume a replacement level of 80 percent of foregone gross wages and for the post-reform period, we assume a replacement level of 100 percent. Equation (5) has been defined for the entire population of treated—but can be approximated using our dataset and weighting observations using frequency weights. By taking the difference of the private sector employer-provided sick pay sums between pre- and post-reform years, we obtain a total maximum increase in labor costs of €5.2 billion for the two post-reform years. However, considering the fact that only roughly half of the private sector workforce was actually affected by the reform, the estimate of the increase in annual costs instead becomes €1.3 billion. This total increase in labor costs can be decomposed into three components as illustrated by equation (6):

$$\begin{aligned}\Delta SLP = \; & \underbrace{0.2 \times \sum_{i=1}^{N} Y_{i0}W_{i0}}_{\text{Replacement Rate Effect}} + \underbrace{\sum_{i=1}^{N} W_{i0}\left(Y_{i1} - Y_{i0}\right)}_{\text{Sick leave effect}} \\ & + \underbrace{\sum_{i=1}^{N} Y_{i1}\left(W_{i1} - W_{i0}\right)}_{\text{Residual}} \end{aligned} \tag{6}$$

The first component is the automatic effect associated with the increase of the federally mandated sick pay level for the first six weeks from 80 to 100 percent of foregone gross wages. We thus disentangle the direct effect from the indirect 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 replacement rate 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 €0.97 billion per year.[22]

[21] 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.

[22] We, thereby, implicitly assume that employees who worked in firms that voluntarily provided 100 percent

The second component of equation (6) represents the indirect behavioral effect, which was triggered by the reform-induced increase in workplace absences. From Table 3, we infer that the 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 effect of €0.8 billion per year.[23] The General Equilibrium Residual is the third component that is caused by time trends, changes in wages for different groups in the workforce, and changes in the employment structure. In the next subsection below, we provide more evidence on general labor market adjustment effects.

The static reform-induced increase in labor costs from the two components is thus $(0.97 + 0.8) =$€1.77 billion per year.[24] Note that this amount exceeds the total effect of €1.3 billion as calculated from equation (5), i.e., the third component is negative. This already provides some evidence that changes in the private sector employment structure might have dampened the potential increase in labor costs. More details on labor market adjustments are provided below.

We also cross-check the plausibility of the 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.[25] 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 sick leave payments 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, we end up with an increase of 8.2 percent.

How Employers' Might Have Compensated for Increased Labor Costs

First, recall that sick leave payments are solely provided by employers and represent labor costs. Since private sector 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 thought to be the main barrier to job creation in Germany. Various researchers studied the relationship between labor costs and job losses by means of general macroeconomic

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 sick leave payments.

[23] Here, we focus on the same data set that we use to obtain the estimated decrease of one day. If we assume that the increase was 0.9 or 1.1 days, we obtain indirect effects of €0.73 and 0.89 billion per year, respectively.

[24] 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 $(1.1 + 0.99) =$ €2.1 billion per year (German Federal Statistical Office, 1996, 1998).

[25] 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.

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.[26]

However, in Germany, dismissal protection is among the strongest worldwide. The very inflexible German labor market might have triggered other attempts at compensation as well. For example, dismissal protection might not allow employers to lay off workers, but their hiring decisions might have been adversely affected. In Table 9, we investigate whether the extensive labor margin was affected in post-reform years relative to pre-reform years in the different sectors. Therefore, we add employees who were registered as unemployed and were actively seeking employment to the sample. For this exercise, we disregard retirees and voluntary non-employment, such as non-employment because of maternity leave.

[Insert Table 9 about here]

The dependent variable in the first column indicates the transition from gainful employment to unemployment, while the dependent variable in the second column measures transitions from unemployment to work between t_0 and t_1. We run parsimonious models and only control for year and sector fixed effects. The first two rows indicate the transition rate in the treatment and control group, respectively, in post-reform years. As seen, about one percent of the sample became unemployed in the two sectors in post-reform years. Although the point estimate is slightly larger for the private sector, the two coefficients do not differ in a statistically significant way.

While we can think of column (1) as representing the firing decisions of employers, column (2) would represent the hiring decision. Interestingly, we find that the transition rate from unemployment to employment is negative and statistically significant for the private sector in post-reform years. It also differs in a statistically significant way from the transition rate in the other sectors, which is not different from zero. Hence, indeed, we find evidence that—probably due to the strict German dismissal protection—the layoff rate did not change, but the hiring rate seem to have been negatively affected in post-reform years.[27]

Next, we examine how labor dimensions on the intensive margin, in addition to workplace attendance, might have adjusted in the aftermath of the reform. The setup of Table 10 is similar to the setup of Table 8, but the dependent variables and the underlying samples used differ. In Table 10, we again use our standard sample of employees and always run our standard OLS-DID model as outlined in equation 1. Column (1) uses work satisfaction as the dependent variable. Column (2) measures the job turnover rate. This outcome measure indicates whether respondents changed jobs between the beginning of the year prior to the interview and the interview. Column (3) makes use of the number of overtime hours per week as dependent variable. Column (4) uses the gross wage per month as dependent variable.

[26] As compared to 28 million private sector employees, i.e., about 0.2 percent. In this very rough calculation, we ignore any other (general equilibrium) effects that might have been triggered by the reform.

[27] Of course, in both cases we observe equilibria that could be influenced either by employer demand decisions or employee supply decisions or a mix of both. However, in a labor market with high structural unemployment, we believe that the short-term employment-unemployment transitions are mainly driven by employer behavior. It is possible, though, that employees, who were on the margin of dropping out of the labor force voluntarily, remained employed because of higher sick leave benefits. It has also been shown that there are spillover effects between health-related social insurance programs. For example, Staubli (2011) finds that a tightening of the Australian DI eligibility rules reduced DI enrollment, but increased the utilization of the sickness insurance program. In our data, however, we do not find evidence that higher sick leave replacement levels affected the transition rate from employment to leaving the labor force voluntarily or vice versa. In total, only 433 people (0.99 percent) of our full-time employed sample between 25 and 55 left the labor force voluntarily and only 1.17 percent took the other road. The pre-and post-reform transition rates hardly differ.

[Insert Table 10 about here]

Consider Panel A first, where we look at the effects for the full sample. There is no evidence that wages, job satisfaction, or the job mobility rate changed in the private sector in post-reform years, relative to the other sectors and pre-reform years. The coefficients are insignificant and very close to zero in magnitude. However, interestingly, we find a highly significant increase in overtime of about 20 minutes per week and employee.

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. Collective bargaining makes it difficult for employers, if not illegal, to discriminate against (unhealthy or unproductive) employees through reduced wage increases. Politicians usually do not implement laws that target these fields.[28] While we do not claim that the relative increase in overtime 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.

Panels B and C now differentiate by pre-reform sick leave days. The results are also robust to using pre-reform employee health as stratifying measure.[29] Let us first consider employees with more than 20 annual sick leave days in one of the pre-reform years (Panel B). Interestingly, there is no evidence at all that overtime increased for this group of employees—the effect is close to zero and not statistically significant. Also, there is no evidence that this group suffered income losses. Interestingly there is highly significant evidence, in line with theoretical predictions of the employer mandate literature (see Section 2), that the job turnover rate increased for this group of the pre-reform unhealthy.

Looking at Panel C and healthy employees with less than 5 annual pre-reform sick leave days, the mirror picture appears: there is no evidence at all for significant changes in the job turnover rate, wages, or job satisfaction for this group, but we observe highly significant increases in overtime work.

In the context of a labor market with strict dismissal protection and a high degree of collective bargaining, the findings from this subsection can be summarized as follows: First, we do not find evidence that employers reacted to the exogenous shock in labor costs by laying off more workers, as standard general equilibrium models would predict. However, instead and in line with theoretical predictions, we find some evidence that employers' hiring decisions may have been affected negatively. Second, we provide evidence that the generosity expansion of the sickness insurance system led to an indirect redistribution between the unhealthy and the healthy—or the productive and the less productive—workers: unhealthy employees reduced their workplace attendance. Employers reacted to the increase in compensated workplace absences by hiring fewer unhealthy employees and letting healthy employees work more overtime. This shift in work burden is also in line with theoretical considerations if, at the establishment level, a fixed amount of work has to be done, but cannot be accomplished by absent or convalescent employees, while employers face high hiring costs in an rigid labor market.

[28] 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, although Bauer et al. (2007) find that this was not the case, at least not for worker flows. To be on the safe side, we excluded all employees working in small establishments with fewer than 20 employees in the empirical models in Table 10.

[29] The results are available upon request from the authors.

7 Conclusion

This article empirically studies the effects of increasing the level of federally mandated employer-provided sickness benefits. The findings illustrate how social insurance interacts with a labor market that is characterized by Bismarckian corporatism. We show that an increase in federally mandated sick pay causally led to a decrease in employee attendance. We also provide evidence on heterogeneity in the reform effects and the reform's impact on employee health. Moreover, we calculate the magnitude of this positive shock to employer benefit payments and empirically study how the labor market adjusted to it.

Making good on a 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, effective January 1, 1999. As a result, employers were mandated to provide standard wage payments for up to six weeks per illness, without any sick leave benefit caps. Public-sector employees and the self-employed were not affected by the benefit increase.

The first part of the empirical section shows 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 matching 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, composition effects, and sample attrition. Our findings suggest that increasing the generosity of the federally mandated sick pay scheme increased sick leave by at least one day per year and employee among the reform's target group. This represents an increase of between 10 and 15 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, which is in line with other SI elasticity estimates (Johansson and Palme, 2005) as well as micro estimates on the unemployment duration-benefit elasticity (Meyer, 1990).

The second part of this paper sheds more light on the underlying mechanisms of the labor supply reactions. We find a great amount of heterogeneity in response behavior to the policy reform. Quantitatively, the effects are mainly driven by self-confident employees, male employees, and employees in a partnership. However, the strongest behavioral reactions are triggered by the small subgroups of employees who knew that they would lose their job, employees with low health satisfaction, and employees with a disability certificate (each about 5 percent of the sample).

Although we use three different health measures as well as the conventional happiness measure, we fail to provide any evidence for reform-related health or happiness improvements. This also holds true for employees in either bad or good pre-reform health. Taken together, this finding supports a shirking explanation and does not suggest that the reform reduced presenteeism significantly—under the assumption that a significant reduction in presenteeism should lead to significant health improvements. Presenteeism and the spreading of contagious diseases to co-workers and customers is one important economic argument in favor of more generous sick leave policies.

Finally, to our knowledge, we are the first to provide empirical evidence as to how the labor market may have adjusted to the expansion of employer-mandated sick leave benefits. Our calculations suggest that total sick leave payments increased by about €1.8 billion or 8 percent 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, i.e., 0.2 percent of total employment. We cannot detect an immediate job loss effect in our data, which might be due to the tight German dismissal protection. However, we find evidence that the reform might have negatively affected the transition from unemployment to employment for employees in bad health. Also, we obtain empirical evidence suggesting a change in the organization of the workforce since overtime hours increased in the private sector relative to other sectors. Interestingly, we find that these compensating differentials for higher sick pay were mainly carried by healthy employees with few pre-reform absence days. This suggests that expanding the generosity of a social sickness insurance system may lead to an redistribution of work load between healthy and unhealthy employees.

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 employer-mandated 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, federally mandated, sick leave. The policy relevance of this topic is reflected in the *Healthy Families Act* currently introduced before both houses of Congress. However, the effects of generosity expansions may differ depending on the initial level of generosity. Clearly, more research on this topic would be illuminating.

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 (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica 74*(1), 235–267.

Abadie, A. and G. W. Imbens (2011). Bias corrected matching estimators for average treatment effects. *Journal of Business and Economic Statistics 29*(1), 1–11.

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

Aon Consulting (2010). One billion man hours lost to sickies across europe each year. `http://aon.mediaroom.com/index.php?s=43&item=1957`, last accessed at August 17, 2012.

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.

Barmby, T., M. G. Ercolani, and J. G. Treble (2002). Sickness absence: an international comparison. *The Economic Journal 112*(480), F315–F331.

Barmby, T. and M. Larguem (2009). Coughs and sneezes spread diseases: an empirical study of absenteeism and infectious illness. *Journal of Health Economics 28*(5), 1012–1017.

Barmby, T., M. Nolan, and R. Winkelmann (2001). Contracted workdays and absence. *Manchester School 69*(3), 269–75.

Barmby, T., J. Sessions, and J. G. Treble (1994). Absenteeism, efficiency wages and shirking. *Scandinavian Journal of Economics 96*(4), 561–566.

Bauer, T. K., S. Bender, and H. Bonin (2007). Dismissal protection and worker flows in small establishments. *Economica 74*(296), 804–821.

Baum, C. L. (2006). The effects of government-mandated family leave on employer family leave policies. *Contemporary Economic Policy 24*(3), 432–445.

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.

Böcken, J., B. Braun, and J. Landmann (2009). *Gesundheitsmonitor 2009* (1 ed.). Verlag Bertelsmann Stiftung.

Böheim, R. and T. Leoni (2011). Firms' moral hazard in sickness absences. IZA Discussion Papers 6005. `http://ftp.iza.org/dp6005.pdf`, last accessed on April 5, 2012.

Boots, S. W., K. Martinson, and A. Danziger (2009). Employers' Perspectives on San Francisco's Paid Sick Leave Policy. Technical report, The Urban Institute. `http://www.urban.org/publications/411868.html`, last accessed on April 5, 2012.

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

Buchmueller, T. C., J. DiNardo, and R. G. Valletta (2011). The effect of an employer health insurance mandate on health insurance coverage and the demand for labor: Evidence from hawaii. *American Economic Journal: Economic Policy 3*(4), 25–51.

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.

Burkhauser, R. V. and M. C. Daly (2012). Social Security Disability Insurance: time for fundamental change. *Journal of Policy Analysis and Management*. forthcoming.

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 (2006). Disability risk and the value of Disability Insurance. In D. Culter and D. Wise (Eds.), *Health In Older Ages: The Causes and Consequences of Declining Disability Among the Elderly*, Chapter 10, pp. 295–336. University of Chicago Press.

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.

D'Amuri, F. (2011). Monetary incentives vs. monitoring in addressing absenteeism: experimental evidence. Economic Working Papers 787, Bank of Italy. `http://www.bancaditalia.it`, last accessed on February 27, 2012.

De Paola, M. (2010). Absenteeism and peer interaction effects: evidence from an Italian public institute. *The Journal of Socio-Economics 39*(3), 420–428.

Delgado, M. A. and T. J. Kniesner (1997). Count data models with variance of unknown form: an application to a hedonic model of worker absenteeism. *The Review of Economics and Statistics 79*(1), 41–49.

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.

Engellandt, A. and R. T. Riphahn (2005). Temporary contracts and employee effort. *Labor Economics 12*, 281–299.

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.

Fevang, E., S. Markussen, and K. Røed (2011). The sick pay trap. IZA Discussion Papers 5655. `http://ideas.repec.org/p/iza/izadps/dp5655.html`, last accessed on April 5, 2012.

Fischer, G., F. Janik, D. Müller, 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 on April 5, 2012.

Frölich, M., A. Heshmati, and M. Lechner (2004). A microeconometric evaluation of rehabilitation of long-term sickness in Sweden. *Journal of Applied Econometrics 19*(3), 375–396.

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.

Goerke, L. and M. Pannenberg (2012). Trade union membership and sickness absence: Evidence from a sick pay reform. SOEPpapers on Multidisciplinary Panel Data Research 470, DIW Berlin, The German Socio-Economic Panel (SOEP).

Goux, D., E. Maurin, and B. Petrongolo (2012). Worktime regulations and spousal labour supply. Technical report. `http://personal.lse.ac.uk/petrongo/Goux_et_al_March2012.pdf`, last accessed on April 11, 2012; older version also available as IZA DP 5639.

Gruber, J. (1994). The incidence of mandated maternity benefits. *American Economic Review 84*(3), 622–41.

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

Hanson, S. G. and A. Sunderam (2012). The variance of non-parametric treatment effect estimators in the presence of clustering. *Review of Economics and Statistics*. forthcoming.

Hashimoto, M., R. Percy, T. Schoellner, and B. A. Weinberg (2004). The long and short of it: maternity leave coverage and womens labor market outcomes. IZA Discussion Papers 1207, Institute for the Study of Labor (IZA).

Hassink, W. H. and P. Koning (2009). Do financial bonuses reduce employee absenteeism? Evidence from a lottery. *Industrial & Labor Relations Review 62*(3), article 4.

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

Heckman, J. J., H. Ichimura, and P. E. Todd (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies 64*(4), 605–54.

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.

Heymann, J., H. J. Rho, J. Schmitt, and A. Earle (2009). Contagion nation: a comparison of paid sick day policies in 22 countries. Technical Report 2009-19, Center for Economic and Policy Research (CEPR). `http://www.cepr.net/documents/publications/paid-sick-days-2009-05.pdf`, last accessed on April 5, 2012.

Hong, S.-H. (2012). Measuring the effect of Napster on recorded music sales: difference-in-differences estimates under compositional changes. *Journal of Applied Econometrics*. forthcoming.

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.

Ichino, A. and R. T. Riphahn (2005). The effect of employment protection on worker effort. A comparison of absenteeism during and after probation. *Journal of the European Economic Association 3*(1), 120–143.

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

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*(9-10), 1879–1890.

Kolstad, J. T. and A. E. Kowalski (2012). Mandate-based health reform and the labor market: Evidence from the Massachusetts reform. NBER Working Papers 17933.

Kostl, A. R. and M. Mogstad (2012). How financial incentives induce Disability Insurance recipients to return to work. Discussion Papers 685, Research Department of Statistics Norway.

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.

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

Markussen, S., K. Røed, O. J. Røgeberg, and S. Gaure (2011). The anatomy of absenteeism. *Journal of Health Economics 30*(2), 277–292.

Medizinischer Dienst der Krankenversicherung (MDK) (2012). www.mdk.de, last accessed at April 27, 2012.

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 on December 19, 2008.

Meyer, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica 58*(4), 757–82.

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.

Nordberg, M. and K. Røed (2009). Economic incentives, business cycles, and long-term sickness absence. *Industrial Relations 48*(2), 203–230.

Pettersson-Lidbom, P. and P. Skogman Thoursie (2012). Temporary disability insurance and labor supply: evidence from a natural experiment. Technical report. forthcoming.

Pouliakas, K. and N. Theodoropoulos (2011). The effect of variable pay schemes on workplace absenteeism. IZA Discussion Papers 5941. `http://ftp.iza.org/dp5941.pdf`, last accessed on April 5, 2012.

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. Technical report, Zentralverband des Deutschen Handwerks. `http://www.zdh.de`, last accessed on June 19, 2009.

Riphahn, R. T. (2004). Employment protection and effort among German employees. *Economics Letters 85*, 353–357.

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.

Schönberg, U. and J. Ludsteck (2007). Maternity leave legislation, female labor supply, and the family wage gap. IZA Discussion Papers 2699, Institute for the Study of Labor (IZA).

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

Staubli, S. (2011). The impact of stricter criteria for disability insurance on labor force participation. *Journal of Public Economics 95*(9-10), 1223–1235.

Summers, L. H. (1989). Some simple economics of mandated benefits. *American Economic Review 79*(2), 177–183.

Thoursie, P. S. (2011). Reporting sick: are sporting events contagious? *Journal of Applied Econometrics 19*(6), 809–823.

van den Berg, G. J., B. Hofman, and A. Uhlendorff (2012). The role of sickness in the evaluation of job search assistance and sanctions. Technical report. mimeo.

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.

Waldfogel, J. (1999). The impact of the Family and Medical Leave Act. *Journal of Policy Analysis and Management 18*(2), 281–302.

Winkelmann, R. (1999). Wages, firm size and absenteeism. *Applied Economics Letters 6*, 337–341.

Winkelmann, R. (2008). *Econometric Analysis of Count Data* (5 ed.). Springer.

Ziebarth, N. R. (2010). Estimating price elasticities of convalescent care programmes. *The Economic Journal 120*(545), 816–844.

Ziebarth, N. R. and M. Karlsson (2010). A natural experiment on sick pay cuts, sickness absence, and labor costs. *Journal of Public Economics 94*(11-12), 1108–1122.

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

Figure 1: Average Sickness Absence Days for Treatment and Control Group over Time

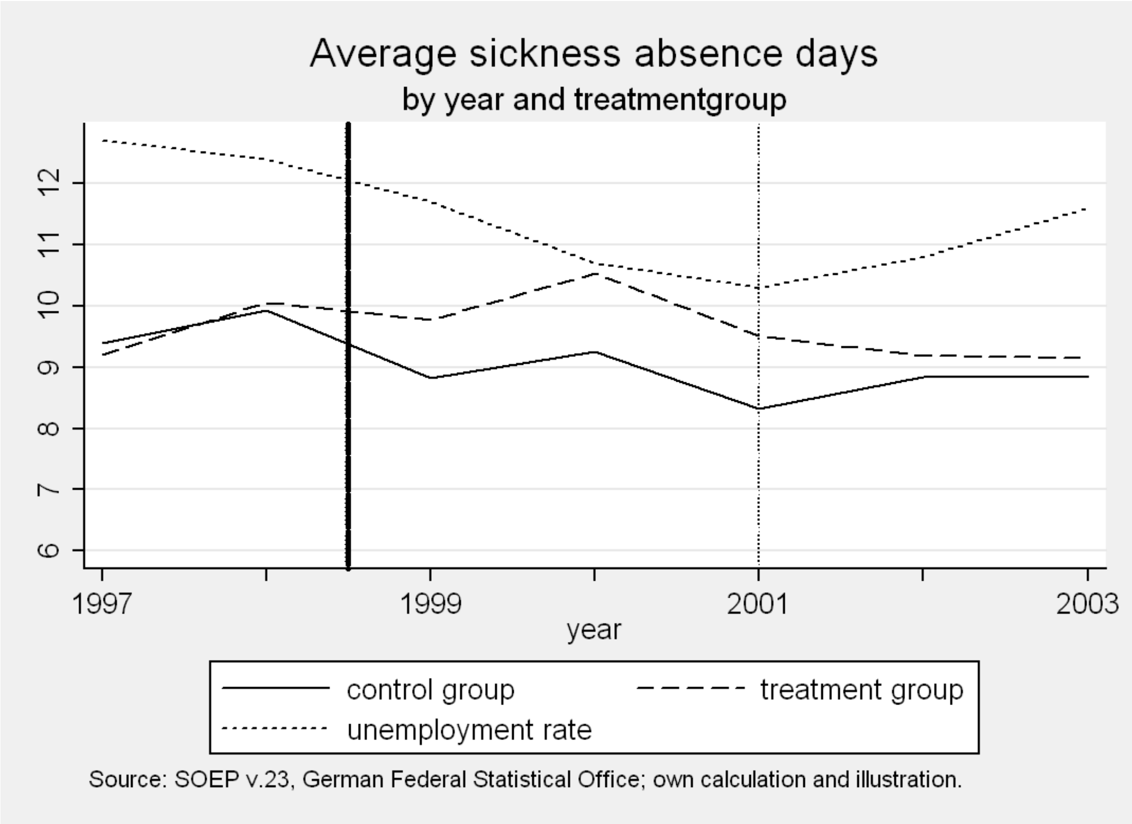

Figure 2: Distribution of Propensity Scores Showing Region of Common Support

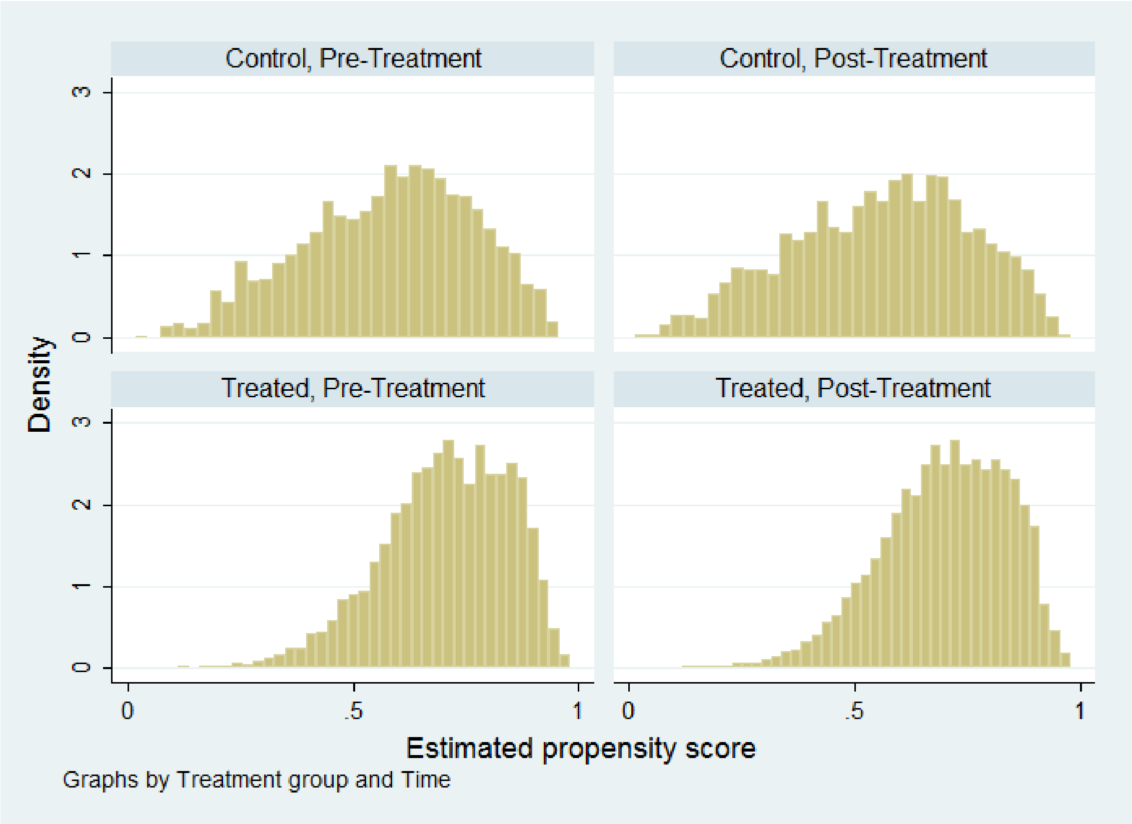

Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Treatment Group	0.6574	0.4746	0	1	22,991
Daysabs	9.7775	24.826	0	365	22,991
A: Variables used in main specifications					
Personal characteristics					
Female	0.3181	0.4657	0	1	22,991
Age	39.7	8.2	25	55	22,991
Age squared	1,644	661	625	3025	22,991
Immigrant	0.1434	0.3505	0	1	22,991
East German	0.265	0.4414	0	1	22,991
Partner	0.7947	0.404	0	1	22,991
Married	0.6807	0.4662	0	1	22,991
Children	0.4794	0.4996	0	1	22,991
Disabled	0.0438	0.2048	0	1	22,991
Health good (best two of five SAH categories)	0.6357	0.4812	0	1	22,991
Health bad (worst two of five SAH categories)	0.0813	0.2734	0	1	22,991
No sports	0.3779	0.4849	0	1	22,991
Educational characteristics					
Drop out	0.0268	0.1616	0	1	22,991
Degree after 8 years' schooling	0.2889	0.4532	0	1	22,991
Degree after 10 years' schooling	0.3654	0.4816	0	1	22,991
Degree after 12 years' schooling	0.0512	0.2203	0	1	22,991
Degree after 13 years' schooling	0.1928	0.3945	0	1	22,991
Other degree	0.0747	0.2629	0	1	22,991
Years with company	9.5221	8.4843	0	41	22,991
Trained for job	0.6064	0.4886	0	1	22,991
Job characteristics					
New job	0.1567	0.3635	0	1	22,991
Blue-collar worker	0.3711	0.4831	0	1	22,991
White-collar worker	0.4604	0.4984	0	1	22,991
One man company	0.033	0.1787	0	1	22,991
Small size company	0.2437	0.4293	0	1	22,991
Medium size company	0.2754	0.4467	0	1	22,991
Large company	0.2209	0.4149	0	1	22,991
Very large company	0.227	0.4189	0	1	22,991
Gross wage per month	2,471	1,293	404	28,632	22,991
Annual state unemployment rate	11.3504	4.5326	5.4	21.7	22,991
B: Variables used in extended analyses					
Long-term absence	0.0559	0.2298	0	1	22,991
Job change previous year	0.1096	0.3124	0	1	22,991
Not impaired by health	0.7829	0.4123	0	1	22,991
Severely impaired by health	0.0237	0.152	0	1	22,991
Low health satisfaction (0-4 on scale up to 10)	0.0528	0.2236	0	1	22,991
High health satisfaction (10 on scale up to 10)	0.0935	0.2912	0	1	22,991
Overtime hours per week	2.6834	3.8912	0	23.1	20,670
High job satisfaction (10 on scale up to 10)	0.0929	0.2904	0	1	22,991

Continued on next page...

... Table 1 continued

Variable	Mean	Std. Dev.	Min.	Max.	N
Low life satisfaction (0-4 on scale up to 10)	0.0285	0.1664	0	1	22,991
No work council('01; no job changers)	0.0559	0.2797	0	1	16,130
Workers laid off last yr.	0.2797	0.4489	0	1	17,150
Expects job loss within 2 years ('99; no job changers)	0.0354	0.1848	0	1	15,911
Expects promotion within 2 years ('98; no job changers)	0.1959	0.3969	0	1	16,779
Not religious ('97)	0.365	0.4814	0	1	15,484
Sickness should be insured by state ('97)	0.3803	0.4855	0	1	15,484
Sickness should be insured privately('97)	0.0855	0.2796	0	1	15,484
Control life ('99)	0.3822	0.4859	0	1	17,291
Can influence life ('99)	0.4087	0.4916	0	1	17,291
Need to work hard for success ('99)	0.5261	0.4993	0	1	17,291

Source: SOEP v.23, own calculation and illustration. 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 *expects job loss within 2 years* question in 1999 are kept in all years in which they were interviewed and worked at the same workplace as in 1999. 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			***Blocked Sample***			***Matched Sample***		
Covariates	**Treat. group**	**Control group**	**Norm. diff.**	**Treat. group**	**Control group**	**Norm. diff.**	**Treat. group**	**Control group**	**Norm. diff.**
Age	39.080	40.908	0.158	39.083	40.801	0.149	39.085	40.494	0.122
Female	0.277	0.397	0.182	0.277	0.391	0.172	0.277	0.367	0.136
Partner	0.798	0.788	0.016	0.798	0.790	0.014	0.798	0.792	0.011
Married	0.677	0.688	0.018	0.677	0.687	0.016	0.677	0.683	0.008
Immigrant	0.178	0.077	0.218	0.176	0.077	0.212	0.176	0.089	0.183
Children	0.488	0.462	0.037	0.488	0.464	0.035	0.488	0.462	0.037
Disabled	0.042	0.048	0.021	0.042	0.047	0.019	0.042	0.048	0.022
Health good	0.633	0.641	0.012	0.633	0.641	0.013	0.633	0.642	0.014
Health bad	0.081	0.083	0.006	0.081	0.083	0.007	0.081	0.083	0.006
8 years of schooling	0.326	0.217	0.175	0.325	0.220	0.169	0.325	0.254	0.111
10 years of schooling	0.350	0.396	0.067	0.351	0.400	0.073	0.351	0.424	0.106
13 years of schooling	0.151	0.273	0.212	0.152	0.264	0.198	0.152	0.196	0.084
Trained for job	0.565	0.686	0.177	0.566	0.682	0.171	0.566	0.651	0.123
New job	0.177	0.117	0.120	0.176	0.119	0.115	0.176	0.127	0.095
Years with company	8.767	10.971	0.183	8.780	10.846	0.172	8.790	10.389	0.135
White collar	0.494	0.395	0.142	0.493	0.400	0.133	0.493	0.453	0.057
Gross wage/1,000	2,392.8	2,621.4	0.120	2,392.2	2,577.3	0.105	2,392.8	2,438.9	0.028
State unemployment rate	11.147	11.740	0.092	11.154	11.757	0.094	11.151	11.816	0.103

Source: SOEP v.23, own calculation and illustration. "Norm. diff." stands for "Normalized difference" which is calculated according to $\frac{X_1 - X_0}{\sqrt{\sigma_1^2 + \sigma_0^2}}$, where $\bar{X}_1$ 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. "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. The covariates used 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, 142 observations (0.6%) are not considered since they lie outside the common support which is [0.1180; 0.9761]. The number of blocks is ten; each one containing an equal number of observations. The "matched sample" has been generated by means of five-to-one nearest neighbors matching based on the propensity score. In total, the raw sample contains 22,991 observations, the blocked sample contains 22,849 observations, and the matched sample contains 21,666 observations.

Table 3: Difference-in-Differences Intention-to-Treat Estimation (DID ITT): Regression, Matching, and Combined Methods

	Regression		***Matching***			***Matching + Regression***	
Variable	**OLS**	**ZINB-2**	**blocking**	**nearest neighbors**	**bivariate match**	**blocking + regression**	**n.neighbors + regression**
DID=D×T	1.3766*	1.0992**	2.2814***	2.0230**	2.2390***	1.4869***	1.6324***
	(0.7100)	(0.4670)	(0.7634)	(0.9352)	(0.2575)	(0.4088)	(0.3328)
Covariates employed							
Job characteristics	yes	yes	yes	yes	yes	yes	yes
Educational characteristics	yes	yes	yes	yes	yes	yes	yes
Personal characteristics	yes	yes	yes	yes	yes	yes	yes
Regional unemployment rate	yes	yes	yes	yes	yes	yes	yes
Time dummies	yes	yes	no	no	no	no	no
State dummies	yes	yes	no	no	no	no	no
N	22,991	22,991	22,849	21,666	21,666	22,849	21,666
N DID=D×T=1	8,967	8,967	8,929	8,929	8,929	8,929	8,929

Source: SOEP v.23, own calculation and illustration; * $p<0.1$, ** $p<0.05$, *** $p<0.01$; 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=D×T (=1). ZINB-2 stands for *Zero-Inflated Negative Binominal Model 2*. In column (3), the propensity score (PS) of belonging to the treatment group is estimated, based on a logit model and the same covariates as in Table 2. Based on this PS, the sample is stratified into ten blocks, each with a roughly equal PS for each of the four groups. 142 observations (38 with DID=D×T=1) are outside the region of common support. Then, the block-specific treatment effects—the double difference between groups over time—are weighted by the number of treated to obtain the overall average treatment effect on the treated. Standard errors in columns (3) and (4) are obtained in accordance with Abadie and Imbens (2006), and in column (5) using bootstrapping with 100 replications. In column (4), the treatment effect is obtained by five-to-one nearest neighbors matching. In column (5) we allow for the possibility of compositional changes by conducting a bivariate local linear matching procedure (Heckman et al., 1997, 1998). In columns (4) and (5), 1,325 (38 with DID=D×T=1) observations are lost due to a lack of overlap in the support of the PS. In column (6), 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 (DID=D×T=1) to obtain the overall treatment effect. In column (7), five-to-one nearest neighbors matching and regression are combined.

Table 4: Difference-in-Differences Full Compliance Estimation (DID ATT): Specific Samples

	Full Sample				***Private Sector Employees***			
	Intended		**Implemented**		**Intended**		**Implemented**	
Variable	**OLS (1)**	**FE (2)**	**OLS (3)**	**FE (4)**	**OLS (5)**	**FE (6)**	**OLS (7)**	**FE (8)**
$DID{=}D{\times}T$	1.7646***	3.2175***	2.0835***	4.0090***	1.6163*	2.4284	1.9760	5.6029**
	(0.3198)	(0.8046)	(0.3016)	(0.8587)	(0.7661)	(3.4051)	(1.5669)	(2.2709)
Covariates employed								
Job characteristics	yes	yes	yes	yes	yes	yes	yes	yes
Educational characteristics	yes	yes	yes	yes	yes	yes	yes	yes
Personal characteristics	yes	yes	yes	yes	yes	yes	yes	yes
Regional unemployment 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
N	11,888	11,888	11,888	11,888	4,011	4,011	4,011	4,011

Source: SOEP v.23, own calculation and illustration; * p<0.1, ** p<0.05, *** p<0.01; standard errors are in parentheses and clustered on the (treatment indicator)×year level. All four 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 (3), (4) and (7), (8) 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 the last four columns, the first four columns include public sector employees and the self-employed, who are codified as being fully treated throughout. The last four columns solely rely on private sector employees. The models labeled “intended” are for the purpose of comparison and assign respondents to the treatment and control group using the same criteria as in the previous analysis.

Table 5: Robustness Checks

Panel A:	**Fixed-Effects**	**weighted no attrition**	**1997 vs. 2000**	**+ lagged daysabs**	**impact on long-term absenteeism**
OLS	2.0434***	1.8859**	1.4327	1.6222**	-0.0034
	(0.7407)	(0.8178)	(0.8936)	(0.7162)	(0.0064)
N	22,991	22,991	11,606	19,164	22,991

Panel B:	**no job change (treated)**	**no post-reform industry changers**	**no sector changers & weighted**	**clustered at state × year level**	**clustered at industry × year level**
OLS	1.3616*	1.3145*	2.1368**	1.3766**	1.3766**
	(0.7545)	(0.7728)	(0.8948)	(0.6055)	(0.6643)
N	18,233	19,076	20,892	22,991	22,991

Source: SOEP v.23, own calculation and illustration; * p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers, except for Panel B column (4), where they are clustered on state×year (64 cluster) and column (5), 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 Panel A, column (1), estimates a Fixed-Effects model. The model in column (2) of Panel A weights the regression with the inverse probability of no dropping out of the sample in post-reform years. Column (3) of Panel A excludes any anticipation or transition effects and compares the years 1997 to the year 2000. Column (4) of Panel A contains the lagged number of annual absence days as an additional covariate. Column (5) estimates the reform effect on the incidence of long-term absenteeism, i.e., a sickness period of more than six weeks. The model in column (1) of Panel B excludes all respondents in the treatment group, who changed their jobs at least once in the period under consideration. Column (2) excludes all those who have changed their industry branch in the post-reform period. Column (3) of Panel B excludes individuals who switched between sectors, i.e., the private sector, the public sector, or self-employment, and additionally weights the regression with the inverse probability of not dropping out of the sample in post-reform years.

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

Model	**1998**	**2000**	**2001**
OLS	-0.5467	0.5192	-0.4435
	(0.8185)	(0.6130)	(0.6272)
ZINB-2	-0.5174	0.7315	-0.4521
	(0.7905)	(0.5482)	(0.5902)
nearest neighbors	-0.3789	0.3468	-0.0617
+ regression	(0.6581)	(0.4459)	(0.4129)
N	16,464	25,692	27,912

Source: SOEP v.23, own calculation and illustration; * p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. All columns make use of two pseudo pre- and two pseudo post-reform years, i.e., column (1) includes the waves 1997-2000, column (2) includes the waves 1999-2002, and column (3) includes the waves 2000-2003. While the coefficients shown estimate the pseudo-reform effect for two pseudo post-reform years in columns (2) and (3), column (1) just identifies the single pseudo reform year 1998, since 1999 was the first true reform year. 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=D×T (=1). Every cell stands for one model.

Table 7: Assessing Heterogeneity in Reform Effects

Panel A: Personal characteristics						
	female	**over 40**	**gross wage > median**	**partner**	**children**	**highest school degree**
$D{\times}T{\times}$[column]	-0.8464	-0.0657	-1.4093*	1.6963*	-0.4537	-0.3609
	(0.8494)	(0.7400)	(0.7759)	(1.0161)	(0.8161)	(0.7319)
Covariate [column]	1.6485	0.7083	-0.7747	-1.5548**	0.2205	-4.6037***
	(1.0458)	(0.4967)	(0.5159)	(0.7449)	(0.5189)	(1.7228)
Covariate [column] = 1|treated	0.2768	0.4236	0.4731	0.7979	0.4883	0.1513
Panel B: Health status						
	health very good	**health bad or poor**	**high health satisfaction**	**low health satisfaction**	**not impaired by health**	**disabled**
$D{\times}T{\times}$[column]	-0.0151	4.6713*	-1.3334*	5.8015	-1.4004	9.5639**
	(0.5905)	(2.6854)	(0.6959)	(3.753)	(1.2398)	(4.1413)
Covariate [column]	-1.706***	7.7736***	-0.9804**	9.1515***	-4.1658***	5.1737***
	(0.3852)	(1.5108)	(0.4742)	(1.9346)	(0.7613)	(1.9228)
Covariate [column] = 1|treated	0.1170	0.0806	0.0955	0.0512	0.7832	0.0417

Source: SOEP v.23, own calculation and illustration; * 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 in Panel A and B include an set of (triple) interaction terms between $DID{=}D{\times}T$ and all covariates displayed in the six column headers in columns (1) to (6) of Panel A, in addition to the respective covariates in levels. However, we only display the triple interaction term with the covariate shown 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. *Children* is a dummy variable that indicates whether the employee has kids. *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* is 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 22,991 observations. The descriptive statistics for all column-header variables used are shown in Table 1.

Panel C: Workplace characteristics

	small firm	**no work council ('01)**	**blue collar worker**	**workers laid off last yr.**	**job loss very likely within 2 yrs. ('99)**	**promotion likely within 2 yrs. ('98)**
$D \times T \times$[column]	1.3769*	1.1222*	-0.0227	-1.1941	8.6155**	-0.4585
	(0.7621)	(0.5818)	(0.7778)	(0.9693)	(3.9169)	(0.5044)
Covariate [column]	1.4733*	-1.5280***	2.6624***	1.5968***	2.1172	-0.4589
	(0.8976)	(0.5818)	(0.7096)	(0.5507)	(1.4680)	(0.5045)
Covariate [column] = 1|treated	0.2522	0.4027	0.5056	0.2823	0.0416	0.1951

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**
$D \times T \times$[column]	-0.7554	0.2130	3.4438	2.1684**	2.2345**	1.6925**
	(1.0508)	(1.0073)	(2.2164)	(0.8978)	(1.0338)	(0.8676)
Covariate [column]	1.6752**	-0.0828	0.1689	-0.8228	-0.3841	0.2777
	(0.7004)	(0.5741)	(0.7896)	(0.4818)	(0.5411)	(0.2169)
Covariate [column] = 1|treated	0.3508	0.4056	0.0801	0.4128	0.2844	0.5398

Source: SOEP v.23, own calculation and illustration; * 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 triple interaction term between $DID = D \times T$ 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 1998 that they expect a job promotion within the next two years are kept in all years in which they had the same workplace as in 1998. *Small firm* in column (1) of Panel C stands for firms with less than 20 workers (22,991 obs.). *No work council* in column (2) selects on repondents who did not change their workplace and indicated, in 2001, that there would not exist a workcouncil at their workplace (16,130 obs.). *Blue collar worker* has a one for blue collar workers (22,991 obs. each). *Workers laid off last yr.* was only sampled in 1998 and 1999 and has a one for those who answered "reduced" to the question "At the workplace where you work now, has the workforce been increased, reduced, or did it remain stable in the last 12 months?" (17,150 obs.). In 1999, respondents were asked on a 11-category scale: "How likely is it that you will loose your job in the next 2 years?" Those who indicated 80, 90, or 100 percent have a one on the dummy variable *job loss expected within 2 yrs. ('99)* in column (5) (15,911 obs.). In column (6), we make use of a very similar question on job promotion which was asked in 1998. Those who found it "very likely" or "likely" that they would get promoted within the next two years have a one on their dummy variable *promotion likely within 2 yrs. ('98)* (16,779 obs.). The variables used in columns (1) to (3) of Panel D were only sampled in 1997 (15,484 obs.). The variables used in columns (4) to (6) of Panel D were only sampled in 1999 (17,291 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 Effects on Employees' Health Status: By Pre-Reform Employee Health Status

Panel A: Full Sample				
	health bad or poor	**sev. impaired by health**	**low health satisfaction**	**low life satisfaction**
DID=D×T	0.0025	-0.0006	0.0005	0.0075
	(0.0077)	(0.0045)	(0.0066)	(0.0048)
Panel B: More than 20 pre-reform absence days				
	health bad or poor	**sev. impaired by health**	**low health satisfaction**	**low life satisfaction**
DID=D×T	0.0112	0.0029	0.0238	0.0079
	(0.0344)	(0.0249)	(0.0301)	(0.0217)
Panel C: Less than 5 pre-reform absence days				
	health bad or poor	**sev. impaired by health**	**low health satisfaction**	**low life satisfaction**
DID=D×T	-0.0019	-0.0021	-0.0037	0.0036
	(0.0094)	(0.0044)	(0.0086)	(0.0061)

Source: SOEP v.23, own calculation and illustration; * p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. Each cell represents one model as in equation (1), estimated by OLS, but uses the corresponding variable in the column header as outcome measure. All outcome variables used are detailed in Table 1. Panel A uses the full sample. Panel B uses a balanced panel and then selects on respondents with more than 20 annual sick leave days in at least one of the pre-reform years. Panel C also uses a balanced panel, but selects on respondents with less than 5 annual sick leave days in at least one of the pre-reform years. The models in Panel A make use of 22,991 person-year observations. The models in Panel B make use of 1,744 person-year observations and the models in Panel C make use of 8,884 person-year observations.

Table 9: Labor Force Transitions: Private vs. Other Sectors

Variable	**Work→ Unemployment**	**Unemployment→ Work**
Private Sector×post-reform	-0.0129***	-0.0089***
	(0.0032)	(0.0033)
Other Sectors×post-reform	-0.0098***	-0.0010
	(0.0036)	(0.0038)
Private Sector	0.0309***	0.0236***
	(0.0025)	(0.0026)
Other Sectors	0.0167***	0.0077***
	(0.0028)	(0.0029)
t1998	-0.0039*	0.0066***
	(0.0020)	(0.0021)
t1999	-0.0040	-0.0082***
	(0.0029)	(0.0030)
t2000	0.0026	-0.0013
	(0.0029)	(0.0030)
N	43,678	43,678
F-test (p-value): Private Sector×post-reform = Other Sectors×post-reform	0.3225	0.015

Source: SOEP v.23, own calculation and illustration; * p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. All columns make use of two pre- and two post-reform years, i.e., years 1997 to 2000. In contrast to all other tables, this sample includes non-working, unemployed, and part-time employed people. However, we exclude retirees, people who are non-working because they are on maternity leave, as well as non-working people in educational training and those who only work from time to time. The dependent variable in column (1) indicates transitions from regular employment in $t-1$ to unemployment in t. Consequently, the dummy variable *Private Sector* indicates whether the respondent was employed in the private sector in $t-1$. The dummy *Other Sectors* has a one for respondents in the public sector or in self-employment in $t-1$. The dependent variable in column (2) indicates transitions from unemployment in $t-1$ to regular employment in t. Consequently, the dummy variable *Private Sector* indicates whether the respondent is employed in the private sector in t. The dummy *Other Sectors* has a one for respondents in the public sector or in self-employment in t.

Table 10: Labor Market Adjustments and Employer Behavior: By Pre-Reform Employee Health Status

Panel A: Full Sample				
	high work satisfaction	**job change prev. year**	**overtime (hrs./week)**	**gross wage (per month)**
DID=*D*×*T*	-0.0068	0.0049	0.3193***	-21.82
	(0.0106)	(0.0091)	(0.1211)	(26.87)
Panel B: More than 20 pre-reform absence days				
	high work satisfaction	**job change prev. year**	**overtime (hrs./week)**	**gross wage (per month)**
DID=*D*×*T*	-0.0099	0.0701***	-0.003	-23.26
	(0.0196)	(0.0253)	(0.3105)	(35.81)
Panel C: Less than 5 pre-reform absence days				
	high work satisfaction	**job change prev. year**	**overtime (hrs./week)**	**gross wage (per month)**
DID=*D*×*T*	0.0157	0.0017	0.3686**	-22.13
	(0.0124)	(0.0118)	(0.1567)	(28.44)

Source: SOEP v.23, own calculation and illustration; * p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. Each cell represents one model as in equation (1), estimated by OLS, but uses the corresponding variable in the column header as outcome measure. All outcome variables used are detailed in Table 1. All models in all panels exlude employees in small establishments with fewer than 20 employees since the dismissal protection exemption threshold was raised from 5 to 10 employees in 1999 (Bauer et al., 2007). Apart from excluding employees in small firms, Panel A uses the full sample. Panel B uses a balanced panel and then selects on respondents with more than 20 annual sick leave days in at least one of the pre-reform years. Panel C also uses a balanced panel, but selects on respondents with less than 5 annual sick leave days in at least one of the pre-reform years. The models in Panel A make use of 15,675 person-year observations. The models in Panel B make use of 1,394 person-year observations and the models in Panel C make use of 6,226 person-year observations.