Sickness Absence and Economic Incentives

Sickness Absence and Economic Incentives

genehmigte Dissertation

von der Fakultät VII - Wirtschaft und Management
der Technischen Universität Berlin (TU Berlin)
zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften
Dr. rer. oec.

von

Nicolas Robert Ziebarth

geboren am 25. Mai 1982
in Frankfurt am Main

Tag der wissenschaftlichen Aussprache: 28. Februar 2011

Berlin 2011
D 83

Doktorvater und 1. Gutachter:
Univ.-Prof. Dr. rer. oec. Gert G. Wagner
Fachgebiet für Empirische Wirtschaftsforschung und Wirtschaftspolitik
Technische Universität Berlin (TU Berlin)
Fakultät VII: Wirtschaft und Management

Vorstandsvorsitzender des Deutschen Instituts für Wirtschaftsforschung (DIW Berlin)

Erste Dienstadresse:
DIW Berlin
Mohrenstrasse 58
D-10117 Berlin
Telefon: +49-(0)30-89789-290
Fax: +49-(0)30-89789-109
E-Mail: gwagner@diw.de
Homepage: http://www.diw.de/en/cv/gwagner

2. Gutachterin (extern):
Univ.-Prof. Regina T. Riphahn, Ph.D.
Friedrich-Alexander Universität Erlangen-Nürnberg
Fachbereich Wirtschaftswissenschaften
Institut für Arbeitsmarkt- und Sozialökonomik
Lehrstuhl für Statistik und empirische Wirtschaftsforschung

Dienstadresse:
Friedrich-Alexander Universität Erlangen-Nürnberg
Lehrstuhl für Statistik und empirische Wirtschaftsforschung
Lange Gasse 20
D-90403 Nürnberg
Telefon: +49-(0)911-5302-268
Fax: +49-(0)911-5302-178
E-Mail: Regina.Riphahn@wiso.uni-erlangen.de
Homepage: http://www.lsw.wiso.uni-erlangen.de

Doktorand:
Dipl.-Vw., Dipl.-Kfm. Nicolas Robert Ziebarth
Promotionsstudent an der TU Berlin
Immatrikulationsnummer: 22 54 73

Mitglied der SOEP-Gruppe am DIW Berlin
Mitglied des Graduate Center of Economic and Social Research am DIW Berlin

Dienstadresse:
DIW Berlin
SOEP Office
Mohrenstrasse 58
D-10117 Berlin

Telefon:	+49-(0)30-89789-587
Fax:	+49-(0)30-89789-109
E-Mail:	nziebarth@diw.de
Homepage:	http://www.diw.de/cv/en/nziebarth

Contents

List of Tables

List of Figures

Acknowledgements

Numerous people have contributed to the final version of this thesis. They have helped and supported me during a time period of almost exactly four years, which is why I am greatly indebted to every one of them.

Starting with those who have supported and inspired me academically, I would like to begin by expressing my gratitude to my supervisor Prof. Dr. Gert G. Wagner. In every single phase of my dissertation, Prof. Wagner encouraged me to pursue my research ideas and plans. He gave me the academic freedom that I needed to remain motivated. His method of supervising and teaching me was optimal for my academic development and has made me work harder than I could have imagined. At the same time, I really enjoyed carrying out the research, also largely thanks to his approach. I would also like to thank my secondary supervisor, Prof. Regina T. Riphahn, for the many hours she spent on my thesis work, the final report, and for her valuable comments and suggestions.

Continuing with my acknowledgments in chronological order, my gratitude goes to those who selected me to become a member of the first cohort of the *Graduate Center of Economic and Social Research* at the *German Institute for Economic Research (DIW Berlin)*, where I conducted my research. I should add that I would not have applied to this program without the inspiration and encouragement that I received from Joachim R. Frick. In 2004, after having had a frustrating time during my Bachelor's studies – which did not meet my expectations about studying economics – I changed universities for my Master's degree. During my Master's studies at the *Berlin University of Technology (TU Berlin)*, for the first time, I was fascinated by economics. At this opportunity, I would also like to thank all my teachers and fellow students at the TU Berlin, especially Prof. Dr. Jürgen Kromphardt, for bringing the fun back into economics. In 2004, as part of its Master's studies program, the TU Berlin offered a student research project in collaboration with the SOEP department

at DIW Berlin. I participated in this project which was led by Joachim R. Frick and Laura Romeu Gordo. For this project, we were instructed to analyze a topic of our choice in teams of two using ©STATA and the *German Socio-Economic Panel Study (SOEP)*. Although it was very challenging discussing research designs with my fellow student for hours on end, this experience was the best I had during my studies. In fact, this was the first time I realized that doing research might be what I wanted to do in the future. This shaped the future course of my life since without that experience I would probably have not applied for the PhD program at DIW Berlin. Joachim Frick encouraged me to apply and was also a member of the Graduate Program Selection Committee. My project partner is now a very good friend of mine. The project also inspired him to become a PhD student in econometrics.

The first year of the DIW PhD program was extremely helpful for me. I am grateful to everyone who contributed to that year and also made it possible for me to have the opportunity to stay in Washington DC at *DIW DC* for three months.

Prof. Wagner and the SOEP department, the department in which I worked on my dissertation, provided me with a great deal of support. The research environment at the SOEP department is excellent. All the long and short academic discussions directly improved the quality of my research. I thank everyone in the SOEP department who contributed to the final product of my thesis, in one or another way, directly or indirectly. In particular, I would like to thank Silke Anger, Eva M. Berger, Joachim R. Frick, Markus M. Grabka, Martin Kroh, Henning Lohmann, Jürgen Schupp, Ingo Sieber, Tom Siedler, Gert G. Wagner, and Michael Weinhardt for helpful discussions and their valuable comments on my work. Very special thanks go to Deborah Bowen who spent many hours copy-editing my papers.

In addition to an excellent academic environment, the financial and infrastructural support available offers great opportunities for young researchers. During the first year of the PhD program, I was funded by a scholarship through DIW Berlin. In Germany, twelve foundations provide selected PhD students with a full-time scholarship that is disbursed for up to three years. The money comes from the Federal Ministry of Education and Research and thus all taxpayers. Thanks to the selection committee of the *Foundation of German Business (Stiftung der Deutschen Wirtschaft, sdw)* and the full-time scholarship that I received from the second to the fourth year of my PhD studies, I was in the fortunate position of being able to spend almost 100 percent of my working hours on my research.

Without the generous funding of the SOEP department, I would not have been able to present my work at dozens of international conferences. Without presenting my work at these conferences, I would not have met all the senior and junior researchers whose comments and discussions were so valuable for the quality of this thesis. Moreover, without the broad scientific network of the SOEP department, I would not have met the co-author of chapters one and two of this dissertation, Martin Karlsson who was a visiting fellow at the SOEP department for three months in spring 2008 (I am greatly indebted to Martin Karlsson for everything that he has done for me. Thanks Martin!). Lastly, without the excellent public reputation that the SOEP group and the DIW have, it would not have been possible to make my research results accessible to a wider public audience. This has been done by numerous newspaper articles which were the outcome of DIW and SOEP press releases and their existing media network. It is extremely motivating and inspiring for a young applied researcher to see that the non-academic public perceives and discusses one's own research results.

During the four years of my PhD studies, I have learnt that one of the most important things to do is to present my work in seminars, at workshops, and at international conferences. It is not only about learning how to present one's own work in a clear and concise manner. It is not only about the comments received after or during a presentation in front of five other scholars, including three other presenters. It is also about meeting other researchers in general. And it is about discussing one's research design with the only person in this five-people session who is eager to hear about your results. Perhaps he or she will invite you to a seminar within the next few months. Or perhaps you will meet ten other researchers working in the same field at this seminar. I have had many similar experiences. I have never regretted attending any conference or workshop and I would always advise everyone to take advantage of these opportunities whenever possible.

I cannot list all the academic scholars I have met during the four years of my studies since they are too numerous to mention. Running the risk of omitting people who were really helpful to me, I will now list academics who have contributed to the development of my thesis through their discussions and comments: all members of the SOEP group, anonymous referees of The Economic Journal, The Journal of Public Economics, The Journal of the European Economic Association, and The American Economic Journal: Applied Economics,

Daniela Andrén, Tim Barmby, Mattias Bokenblom, Christian Boehler, Jörg Breitung, Amitabh Chandra, Laurens Cherchye, Meltem Daysal, Stefano DellaVigna, Liran Einav, Eberhard Feess, Joachim R. Frick, Christina Gathmann, Murat Genç, David Granlund, John P. Haisken-DeNew, Daniel S. Hamermesh, Barbara Hanel, Lars Hultkrantz, Guido Imbens, Per Johansson, Jochen Kluve, Stephen Knowles, Martin Karlsson, Mathias Kifmann, Tobias Klein, Michael Kvasnicka, Sonja Kassenböhmer, Michael Lechner, Henning Lohmann, Steve Machin, Olivier Marie, Bruce D. Meyer, Raymond Montizaan, Andrew Newell, Therese Nilsson, Martin Olsson, Dorian Owen, Mårten Palme, Per Pettersson-Lidbom, Steve Pischke, Nigel Rice, Regina T. Riphahn, Martin Salm, Hendrik Schmitz, John Karl Scholz, Tom Siedler, Peter Sivey, Peter Skogman Thoursie, Jan C. van Ours, Frederic Vermeulen, Tarja Viitanen, Johan Vikström, Roger Wilkins, Gert G. Wagner, and Mark Wooden.

Moreover, I am grateful to all session participants at the following conferences:

2010

- American Economic Association (AEA 2010), Atlanta, USA
- Econometrics of Healthy Human Resources, Applied Econometrics Association, Rome, Italy
- European Economic Association (EEA 2010), Glasgow, UK
- European Society for Population Economics (ESPE 2010), Essen, Germany
- German Association of Health Economists (dggö 2010), Berlin, Germany
- Health, Happiness, Inequality: Modelling the Pathways between Income Inequality and Health, Darmstadt, Germany

2009

- European Association of Labour Economists (EALE 2009), Tallinn, Estonia
- European Economic Association (EEA 2009), Barcelona, Spain
- Econometric Society European Meeting (ESEM 2009), Barcelona, Spain
- European Society for Population Economics (ESPE 2009), Seville, Spain
- International Conference on Panel Data, Bonn, Germany

- Royal Economic Society (RES 2009), Guildford, UK

2008

- European Association of Labour Economists (EALE 2008), Amsterdam, the Netherlands
- European Economic Association (EEA 2008), Milan, Italy
- European Society for Population Economics (ESPE 2008), London, UK
- Latin American and Caribbean Economic Association (LACEA 2008), Rio de Janeiro, Brazil
- Latin American Meeting of the Econometric Society (LAMES 2008), Rio de Janeiro, Brazil

In addition, I thank everyone who participated in the following seminars or workshops:

2010

- Applied Economics and Econometrics Seminar, University of Mannheim, Mannheim, Germany
- Econometrics Seminar, University, Tilburg University, Tilburg, Netherlands
- Joint Empirical Social Science (JESS) Seminar, Institute for Social & Economic Research (ISER), Colchester, UK
- Melbourne Institute Seminar Series, University of Melbourne, Melbourne, Australia
- PhD Presentation Meeting, Royal Economic Society (RES), London, UK
- University of Otago, Department of Economics, Dunedin, New Zeeland

2009

- Berlin Network of Labour Market Researchers (BeNA), Berlin, Germany
- Brown Bag Seminar, Stockholm University, Stockholm, Sweden
- Brown Bag Seminar, Örebro University, Örebro, Sweden
- Health Economists' Study Group (HESG), Manchester, UK
- Scientific Advisory Board Meeting, DIW Berlin, Berlin, Germany

- Seminar in Health, Labour and Family Economics, Lund, Sweden
- Workshop on Absenteeism and Social Insurance, Uppsala, Sweden
- IZA European Summer School in Labor Economics, Buch, Germany

2008

- Berlin Network of Labour Market Researchers (BeNA), Berlin, Germany
- Fourth GSSS-SOEP Symposium, Delmenhorst, Germany
- Marie Curie Training Programme in Applied Health Economics, Coimbra, Portugal
- Masterclass by Prof. Lindeboom & Prof. Mullahy, Coimbra, Portugal
- SOEP Brown Bag Seminar, Berlin, Germany

Academic support and inspiration is a necessary but not the only condition for being able to write a successful PhD thesis without suffering too much during the process. I thank my parents for their constant belief in me. Moreover, I thank all other family members for supporting me. Friends are irreplaceable. Therefore, I am incredibly grateful to all my friends who always had time for a beer during hard times and time for two beers during good times. Last and most importantly, my deepest gratitude goes to Judith. Without her love and support, I would not have been able to write this thesis.

Nicolas R. Ziebarth
February 2011

Introduction

This doctoral thesis deals with sickness absence and economic incentives. It analyzes how economic incentives, as set by policy makers, shape the decision of employees to go on sick leave.

Despite its enormous relevance, the causes and consequences of workplace absences due to sickness are an under-researched field in economics to date. While there are a large number of studies that build upon correlates of sick leave behavior, studies that convincingly identify how incentives causally affect sick leave behavior are scarce.

In Germany, four percent of contracted labor is lost every year due to sickness absence (Badura et al., 2008). According to OECD data, German employees take an average of 16.5 days' sick leave per year, while the average number of days' sick leave varies drastically among OECD countries between 4.1 (US) and 29.2 (Slovakia) (OECD, 2006). Currently, German employers spend about €25 billion per year for statutory employer-provided sick pay. This sum exceeds one percent of the total GDP. In addition, the public health insurance fund pays out about €6 billion annually for the long-term sick (German Federal Statistical Office, 2009b). To obtain a complete picture of the total amount of benefits paid due to work disability, one would need to add spending by private health insurance companies, disability insurance, and accident insurance which replaces income losses due to work-related disability or accidents.

However, simply summing up the total benefits severely underestimates the total economic costs of sickness absence. For example, if the institutional framework is unable to prevent employees on temporary long-term sick leave from becoming permanently disabled, the economy loses a valuable qualified labor force. Especially in times of demographic change and a shrinking workforce, one crucial challenge of the social insurance system is to maintain employees' capacity to work as long as possible.

To draw on another example about how sick leave behavior may trigger indirect but important side effects, think about an inefficient sickness insurance system. On the one hand, particularly when benefit levels are generous, one may suspect shirking behavior of playing a substantial role. In other words, the fraction of employees who go on sick leave despite being able to work may be substantial. On the other hand, especially when benefit levels are less generous or the unemployment rate is high, presenteeism may be of importance. In other words, when a large proportion of employees go to work despite being sick, this may lead to spillover effects at the workplace due to the spreading of diseases.

While most of these underlying causes and individual reasonings are difficult to unravel empirically, this thesis intends to shed some light on various aspects of how sick leave behavior is shaped. I analyze how sick leave behavior is affected by economic incentives that are varied by different policy reforms. At the same time, I attempt to derive conclusions about the importance of the institutional setting for successful implementation of such reforms. Moreover, it is primarily the institutional setting that determines which actor carries the financial burden through the legal obligation to provide benefits. Since employers in Germany are obliged to provide sick pay for up to six weeks, I provide evidence on how policy reforms impact labor costs and how the labor market might adjust to such shocks in labor costs.

The thesis consists of five independent chapters. Each chapter represents one research study and evaluates at least one specific policy reform in Germany. The unifying aspects of all studies, and hence this thesis, are the following. Firstly, each chapter deals with sickness absence behavior and economic incentives. Secondly, each chapter evaluates policy reforms that were implemented in the mid-1990s in Germany. Thirdly, each chapters builds upon the only data set that contains representative sick leave information for the whole of the German population: the Socio-Economic Panel Study (SOEP). Finally, each chapter makes use of the most recent microeconometric evaluation methods which can be classified as "reduced-form" or "non-structural."

Chapter 1 evaluates how a cut in the replacement level of the statutory sickness insurance in 1996 affected the sick leave behavior of private sector employees. Here, I also approximate the impact on labor costs and calculate potential employment effects, one of the main reform objectives.

Chapter 2 evaluates how the reversal of the reform in 1999 affected sick leave behavior and labor costs. In this chapter, I make use of the rich SOEP data set to provide evidence on the underlying driving forces of the behavioral reactions and on heterogeneity in the reform effects. By characterizing the employees who were mainly responsible for the causal reactions, I also provide evidence on the potential significance of shirking behavior and presenteeism. Lastly, this chapter presents empirical evidence on how employers might have reacted to the shock in labor costs in a rigid labor market with strict dismissal protection.

Chapters 1 and 2 also serve as examples of how reform intention and actual reform implementation of labor market reforms may diverge in a labor market that is characterized by Bismarckian corporatism and a high degree of collective bargaining. The organizational structure of such labor markets restricts policy makers to merely setting federal minimum standards. However, they have no control over the actual reform enforcement on the firm level, which enhances the risk of unpopular reforms failing.

Chapter 3 evaluates how a cut in statutory long-term sick pay affected the sick leave behavior of the long-term sick. The underlying causes of long-term sickness differ substantially from those of short-term sickness. The former are dominated by severe illnesses such as cancer or mental illnesses. The latter are mostly driven by minor diseases such as flu. Thus, a priori, one would suspect that employees on long-term sick leave react differently to economic incentives than those on short-term sick leave. The degree of behavioral adaption to changes in the institutional parameters also sheds light on the importance of moral hazard in the sick leave insurance systems for both long-term and short-term illnesses.

Chapter 4 deals with convalescent care treatments. Convalescent care therapies at health spas involve periods of at least three weeks of workplace absence for employees. I analyze how price responsive the demand for this type of medical care is by evaluating a reform that doubled the daily copayments for these treatments. I also derive price elasticities of demand for specific types of convalescent care therapies.

The last chapter shows how effective direct policy measures are in comparison to indirect measures. Direct policy measures such as increasing copayments unambiguously and universally affect the target population. Indirect measures such as cutting statutory sick pay or widening the options for employers to cut paid vacation in the event of long sick leave simply decrease social minimum standards.

Each chapter is independent and may stand alone. However, as explained above, they are all intrinsically related to one another, methodologically as well as with regard to their underlying content.

General Abstract

Evaluating the causal effects of various policy reforms in Germany, this thesis analyzes how economic incentives and sick leave behavior interact. The analyses have been carried out using microeconometric methods and household survey data of the German Socio-Economic Panel Study (SOEP). The main findings can be summarized as follows:

Firstly, employees clearly adapt their short-term sick leave behavior to changes in sick leave benefits. In 1996, the German legislator cut the statutory sick leave replacement level for private sector employees from 100 to 80 percent of foregone gross wages. This measure led to a fall in the average number of sick leave days of about 12 percent. Reversing the reform in 1999 increased average sick leave days by almost the same amount, namely by 10 percent. When interpreting these figures, one needs to consider that only about half of all private sector employees were effectively affected by both reforms. The response to the reform suggests that moral hazard plays a substantial role in the statutory sickness insurance and in the lower tail of the sickness spell distribution. However, whether the behavioral effects were primarily triggered by shirkers and employees who went on sick leave although they were able to work or by presenteeism and employees who went to work although they were sick remains an open question. I present empirical evidence that is in line with each of the two explanations.

Secondly, changing statutory sick leave benefit levels in an economy that is characterized by Bismarckian corporatism may lead to unintended side effects. Since policy makers can only vary federal minimum standards, their influence on how reforms are enforced on the firm level is limited. Employers are free to provide fringe benefits over and above statutory minimum standards. In the course of the first reform that reduced sick pay, it became clear that German society values a high level of statutory sick leave. A replacement level that falls behind what employees usually

earn is considered to be unsocial by the majority of the Germans. The unpopular cut in sick pay provoked strikes and mass demonstrations. Eventually, many employers agreed in collective wage agreements upon a voluntary provision of 100 percent sick leave. Estimates suggest that between 1997 and 1999 only half of all private sector employees effectively experienced a decrease in actual sick pay. In this particular reform, the divergence between reform intentions as envisioned by the policy makers and actual reform implementation was substantial. In 1999, large parts of the reform were reversed.

Thirdly, changes in the sick pay levels directly translate into changes in labor costs. This is because, in Germany, employers alone are legally obliged to pay for statutory sick leave. My calculations suggest that the changes in the level of statutory sick pay by 20 percentage points represented labor cost changes of approximately €1.5 billion per year, considering that only half of all employees' sick pay was effectively cut. In a completely flexible labor market, economic theory would suggest that these exogenous shocks in labor costs would directly and immediately translate into changes in employment. Relating the estimated changes in labor costs to the findings of other studies for Germany that used general equilibrium models to estimate the relationship between labor costs and employment, the employment dimension of the sick leave reforms equates to around 50,000 jobs, relative to 30 million employees. However, since employment protection is high in Germany, employers were likely to react to the shocks in labor costs in other ways. I present empirical evidence suggesting that in the aftermath of the increase in statutory sick pay, overtime hours increased and wages decreased in the private sector relative to other sectors.

Fourthly, I did not find any evidence that long-term sickness absence of over six weeks is characterized by monetary considerations. In other words, the long-term sick do not adapt their sick leave behavior to economic incentives. Evaluating the effects of a cut in statutory long-term sick leave from 80 to 70 percent of foregone gross wages in 1997, I did not find any empirical evidence for behavioral reactions. A theoretical model confirms this finding under the assumption that the long-term sick are seriously sick. Thus, moral hazard seems to be less of an issue in the upper tail of the sickness spell distribution. One explanation for this finding is the moderateness of the cut in long-term sick pay, which represented on average seven percent of the net wage. Another explanation is linked to the severity of the underlying illness. While short-term sick leave is mainly determined by flu and minor ailments, the main catalysts of long-term sickness absence are cancer, chronic diseases, or mental illnesses. It is plausible that employees diagnosed with cancer, or who are mentally ill, are not very responsive to economic incentives.

Fifthly, I show that the demand for convalescent care treatments is price responsive. For employees, convalescent care therapies at health spas entail sick leave periods of at least three weeks. The decision by the government to double the daily copayments for convalescent care treatments from 1997 onwards induced a decrease in the incidence of these therapies of about 20 percent. Since the question of the price elasticity of demand for medical care is central to health economics, I derive the following price elasticities: the price elasticity for medical rehabilitation therapies to avoid permanent work disability is inelastic and lies around -0.3. Slightly smaller but still inelastic is the price elasticity for medical rehabilitation therapies to recover from accidents at work. According to my calculations, it is about -0.5. Conversely, I find the price elasticity for preventive therapies at health spas to be less than -1 and thus elastic.

Finally, I show that while increasing copayments was very effective in dampening the demand for convalescent care, other cost containment instruments were less effective. I did not find any evidence that a newly introduced option for employers to cut paid vacation in case of sickness absence due to convalescent care was effective. There is no empirical evidence for the notion that this measure reduced the demand for convalescent care therapies at health spas. Likewise, the cut in statutory sick pay did not reduce the incidence of convalescent care therapies. I have two main explanations for these last two findings. First, the cut in paid vacation may have had no effect since many employees use some or all of their vacation days for convalescent care in any case. Although entitled to take paid sick leave in addition to their paid vacation, many employees fear negative job consequences, especially when unemployment rates are high. Second, the cut in sick pay was not necessarily a binding constraint for most employees since they might have faced a decision between going to rehabilitation of simply staying at home to recover. They would have been on sick leave anyway.

All in all, evaluating five different policy reforms and their effects on sick leave behavior, I show that short-term sick leave behavior in particular is responsive to economic incentives. For sickness periods of over six weeks, I did not find such evidence. Policy makers should consider the institutional setting and potential side effects when implementing reforms. When policy makers are in a position to vary parameters that directly and universally affect the target population, the likelihood of successful implementation of the reform is high. As long as policy makers can only vary minimum standards in the regulatory framework of the economy, many actors within the institutional framework may successfully prevent the enforcement of the reform as intended by the policy makers. This is especially true of unpopular reforms.

General Abstract (German)

Diese Arbeit analysiert, wie sich ökonomische Anreize auf das Krankheitsverhalten von Arbeitnehmern auswirken. Dazu werden die Kausaleffekte mehrerer Politikreformen in Deutschland anhand mikroökonometrischer Methoden und Umfragedaten des Sozio-oekonomischen Panels (SOEP) identifiziert. Die zentralen Ergebnisse lassen sich folgendermaßen zusammenfassen:

Erstens wird deutlich, dass Arbeitnehmer kurze Krankheitsepisoden Änderungen im Niveau der Lohnfortzahlung anpassen. Die Kürzung der Lohnfortzahlung im Krankheitsfall von 100 auf 80 Prozent des Bruttolohnes im Jahr 1996 führte dazu, dass die Zahl an Fehltagen von Beschäftigten in der Privatwirtschaft um 12 Prozent pro Jahr sank. Die Rückgängigmachung dieser Reform im Jahr 1999 führte zu einem fast identischen Anstieg der Fehltage um etwa 10 Prozent. Dabei ist zu berücksichtigen, dass schätzungsweise nur die Hälfte aller Arbeitnehmer im privaten Sektor effektiv von den Änderungen betroffen war. Die Reaktion der Arbeitnehmer auf die Änderungen bei der Entgeltfortzahlung zeigen, dass das Phänomen des "moral hazard" im Sozialversicherungssystem der Lohnfortzahlung im Krankheitsfall von Bedeutung ist. Die Verhaltensänderungen könnten einerseits durch Arbeitnehmer, die sich trotz Arbeitsfähigkeit krank melden, verursacht sein. Ein alternativer Erklärungsansatz zielt auf das Phänomen des Präsentismus ab, wonach Arbeitnehmer trotz Krankheit zur Arbeit gehen. Meine empirischen Befunde sind mit beiden Erklärungsansätzen vereinbar.

Zweitens können Änderungen bei der Lohnfortzahlung in einem Wirtschaftssystem, das vom Bismarckschem Korporatismus geprägt ist, unerwünschte Nebeneffekte nach sich ziehen. Da die Politik lediglich Rahmenbedingungen setzen kann, ist ihr Einfluss auf die tatsächliche Umsetzung der beabsichtigten Reformen auf der Firmenebene begrenzt, da Arbeitgeber stets Sozialleistungen oberhalb des gesetzlichen Minimums zahlen können. Im Zuge der Kürzung der Lohnfortzahlung wurde

deutlich, dass die deutsche Gesellschaft eine soziale Absicherung im Krankheitsfall auf hohem Niveau präferiert. Ein Lohnersatzniveau unterhalb des üblichen Gehaltes wird demnach von der Mehrheit der Bevölkerung als unsozial angesehen. Die unpopuläre Reform von 1996 resultierte in etlichen Streiks und Massendemonstrationen. Schließlich stimmten etliche Arbeitgeber in Flächentarifverträgen einer freiwilligen Beibehaltung der alten 100 Prozent Lohnfortzahlungsregelung zu. Es wird geschätzt, dass zwischen 1997 und 1999 lediglich die Hälfte aller Angestellten in der Privatwirtschaft effektiv eine Kürzung der Entgeltfortzahlung hinnehmen musste. Das Auseinanderfallen von Reformabsicht und tatsächlicher Reformumsetzung war dementsprechend groß. Im Jahr 1999 wurden große Teile der Reform rückgängig gemacht.

Drittens wirken sich Änderungen im Niveau der Lohnfortzahlung direkt auf das Niveau der Arbeitskosten aus, da in Deutschland ausschließlich die Arbeitgeber die Lasten der Lohnfortzahlung im Krankheitsfall tragen. Meine Berechnungen ergeben, dass die Änderungen des Lohnersatzniveaus um 20 Prozentpunkte zu Arbeitskostenänderungen von 1,5 Milliarden Euro pro Jahr führten – unter der Annahme, dass nur die Hälfte aller Arbeitnehmer betroffen war. Gemäß der ökonomischen Theorie resultieren in einem völlig flexiblen Arbeitsmarkt exogene Arbeitskostenschwankungen unmittelbar in Beschäftigungseffekten. Unter Zuhilfenahme anderer deutschlandbezogener Studien, die anhand allgemeiner Gleichgewichtsmodelle die Auswirkungen von Änderungen in den Arbeitskosten auf die Beschäftigung schätzen, ergibt sich eine Beschäftigungsdimension der deutschen Reformen in Höhe von etwa 50.000 Arbeitsplätzen, relativ zu gut 30 Millionen Beschäftigten. Aufgrund des strikten Kündigungsschutzes in Deutschland ist es jedoch wahrscheinlich, dass die Arbeitgeber auf andere Art und Weise auf die von den Reformen induzierten Arbeitskostenänderungen reagierten. Diese Arbeit liefert empirische Evidenz dafür, dass in den Folgejahren der Erhöhung der Lohnfortzahlung die Zahl der Überstunden stieg sowie die Löhne - relativ zu anderen Berufsgruppen - sanken.

Viertens finde ich keine empirischen Hinweise darauf, dass Langzeitkrankheit von mehr als sechs Wochen von monetären Überlegungen beeinflusst wird. Mit anderen Worten: Langzeitkranke passen ihr Krankheitsverhalten nicht ökonomischen Anreizen an. Im Jahr 1997 wurde das Krankengeld für gesetzlich Versicherte von 80 auf 70 Prozent des Bruttolohnes gekürzt. Die Evaluierung dieser Reform liefert keine empirische Evidenz auf dadurch ausgelöste Änderungen im Langzeitkrankheitsverhalten. Ein theoretisches Modell bestätigt diesen Befund unter der Annahme, dass Langzeitkranke schwer erkrankt sind. Daher scheint “moral hazard” im Bereich langer Krankheitsdauern von geringer Bedeutung zu sein. Eine Erklärung für diese Erkenntnis könnte auch die Höhe der Kürzung sein. Sie entsprach im Durchschnitt sieben Prozent des Nettolohnes und ist als moderat anzusehen. Ein weiterer Erklärungsansatz zielt auf die Schwere der zugrunde liegenden Krankheit ab. Während kurze Krankheitsdauern vorrangig von Grippe und leichten Erkältungen determiniert werden, sind die Hauptursachen langer Krankheitsdauern Krebs, chronische Krankheiten oder psychische Erkrankungen. Es erscheint plausibel, dass Arbeitnehmer mit einer Krebsdiagnose oder mit psychischen Erkrankungen gegenüber monetären Anreizen nicht sehr empfänglich sind.

Fünftens zeigt diese Arbeit, dass die Nachfrage nach Kuren preissensibel ist. Für Arbeitnehmer bedeuten Kuraufenthalte Arbeitsunfähigkeitsperioden von mindestens drei Wochen. Die Entscheidung der Regierung ab 1997 die täglichen Zuzahlungen für Kuren zu verdoppeln, hat zu einem 20-prozentigen Rückgang von Kuraufenthalten geführt. Da die Frage nach der Preiselastizität der Nachfrage nach Gesundheitsleistungen in der Gesundheitsökonomie zentral ist, konnten die folgenden Preiselastizitäten daraus abgeleitet werden: die Preiselastizität für medizinische Rehabilitationsleistungen zur Vermeidung permanenter Arbeitsunfähigkeit ist unelastisch und liegt bei ungefähr -0.3. Etwas niedriger, aber dennoch unelastisch, ist die Preiselastizität für medizinische Rehabilitationsleistungen zur Überwindung von Unfallfolgen. Nach meinen Berechnungen liegt sie bei etwa -0.5. Im Gegensatz dazu ist die Preiselastizität für präventive Vorsorgekuren kleiner als -1 und daher elastisch.

Abschließend zeigt diese Arbeit, dass die Zuzahlungserhöhung ein sehr effektives Mittel zur Dämpfung der Nachfrage nach Kuren war, wohingegen andere Kostendämpfungsmaßnahmen weniger effektiv waren. Ich finde keinerlei Hinweise darauf, dass eine neu geschaffene Option für Arbeitgeber, im Fall von Kuraufenthalten die Urlaubstage zu kürzen, die Nachfrage nach Kuraufenthalten effektiv reduziert hat. Ebenso wenig hat die Kürzung der Lohnfortzahlung zu einer Reduktion der Kuraufenthalte geführt. Für diese Befunde habe ich zwei Erklärungen. Erstens, die Kürzung des Urlaubsanspruches könnte wirkungslos geblieben sein, da viele Arbeitnehmer ohnehin ihren Urlaub oder Teile ihres Urlaubes für Kuraufenthalte verwenden. Obwohl Arbeitnehmer berechtigt sind, im Falle von Kuraufenthalten zusätzlich zu ihrem Urlaub Lohnfortzahlung im Krankheitsfall zu beziehen, befürchten viele Arbeitnehmer Nachteile im Job, insbesondere bei hoher Arbeitslosigkeit wie Mitte der 1990er Jahre. Zweitens, die Kürzung der Lohnfortzahlung stellte für die meisten Arbeitnehmer nicht unbedingt eine Restriktion dar, denn sie dürften im Wesentlichen eine Entscheidung zwischen einem Kuraufenthalt und der Auskurierung ihrer Krankheit zu Hause getroffen haben. In beiden Fällen wären sie krank geschrieben gewesen.

Anhand der Evaluierung der Effekte von fünf verschiedenen Reformen auf das Krankheitsverhalten von Arbeitnehmern kann diese Arbeit zeigen, dass ökonomische Anreize insbesondere Auswirkungen auf kurze Krankheitsdauern haben. Im Falle von Krankheitsepisoden von mehr als sechs Wochen ist keine solche Evidenz zu finden. Politiker sollten die institutionellen Rahmenbedingungen und potentielle Nebeneffekte berücksichtigen, wenn sie Reformen verabschieden. Wenn Politiker die Möglichkeit haben, Parameter zu ändern, die direkt und umfassend die Zielgruppe ihrer Reformen treffen, ist die Wahrscheinlichkeit einer erfolgreichen Umsetzung des Reformvorhabens hoch. Solange von Politikern nur Rahmenbedingungen und gesetzliche Mindeststandards geändert werden können, haben viele Akteure innerhalb des institutionellen Rahmens die Möglichkeit, eine erfolgreiche Umsetzung der Reform im Sinne der Politiker zu verhindern. Dies gilt insbesondere im Falle unpopulärer Reformen.

Chapter 1

A Natural Experiment on Sick Pay Cuts, Sickness Absence, and Labor Costs

Published in the JOURNAL OF PUBLIC ECONOMICS, 94(11-12): 1108-1122

Abstract

This chapter estimates the reform effects of a reduction in statutory sick pay levels on sickness absence behavior and labor costs. German federal law reduced the legal obligation of German employers to provide 100 percent continued wage pay for up to six weeks per sickness episode. In 1996 statutory sick pay was decreased to 80 percent of foregone gross wages . Within the reform's target group—private sector employees—this measure increased the proportion of employees having zero days of absence between 6 and 8 percent. Quantile regression estimates indicate that employees with up to 5.5 annual absence days reduced their days of absence by about 12 percent. Extended analyses suggest that in industries that enforced the cut, behavioral effects were about twice as large. I show that the direct labor cost savings effect stemming from the cut in replacement levels clearly exceeds the indirect effect due to the decrease in absenteeism. My calculations about the total decrease in labor costs are very much in line with official data which suggest that total employer-provided sick pay decreased by 6.7 percent or €1.7 billion per year.

1.1 Introduction

The relationship between unemployment benefits and unemployment duration has attracted labor economists' attention for decades and provided material for countless numbers of publications. In light of this, it is odd that comparably little research on the relationship between sick leave benefits and sickness absence exists, despite its enormous relevance for labor supply, labor costs, labor productivity, population health, and the functioning of social insurance systems as well as private insurance markets.

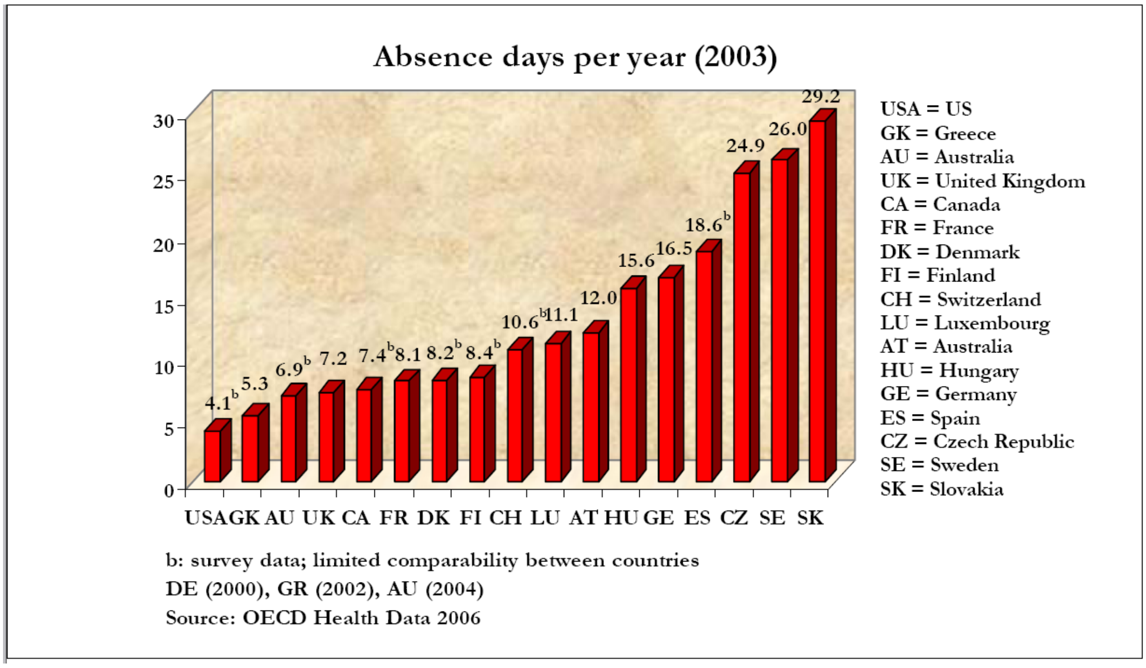
While in Europe ownership of sickness absence insurance is widespread and mostly universal, there is no comparable social insurance at the federal level in the US At the federal level, the US only has disability insurance (DI), which compensates wage losses due to work disability. As compared to the sickness insurance, the literature on the DI is rich (cf. Bound (1989); Gruber (2000); Campolieti (2004); Chen and van der Klaauw (2008); de Jong et al. (2010)). However, the empirical findings on behavioral reactions towards generosity expansions or contractions of the DI are mixed. In addition, these results are unlikely to be transferable to the system of the European sickness insurance since the DI is about permanent rather than temporary withdrawals from the labor market and hence focuses on labor supply at the extensive rather than the intensive margin. The US also knows another social insurance which is run on a state-by-state basis: the Workers' Compensation Insurance (WCI) solely covers income losses due to work-*related* injury or sickness.

Very few studies explicitly analyzed the impact of sick pay levels on absence rates. A handful of studies exploit legislative changes in the benefit levels in Sweden (Johansson and Palme, 1996, 2002, 2005; Henrekson and Persson, 2004; Pettersson-Lidbom and Skogman Thoursie, 2008). Two English studies provide some correlation-based evidence using 1970s era data from (Doherty, 1979; Fenn, 1981). In addition, two papers analyze the impact of changes to WCI benefit levels in the US (Curington, 1994; Meyer et al., 1995). All of the aforementioned studies find that employees adapt their absence behavior to increases and decreases in benefit levels. This finding is reinforced by various other empirical studies which analyze further determinants of sickness absence behavior. Workplace conditions are relevant (Dionne and Dostie, 2007) as are probation periods and economic upswings or downturns (Ichino and Riphahn, 2005; Askildsen et al., 2005).

Average sickness absence days differ substantially across countries, ranging from

4 to 29 days per year and employee (see Figure 1.1). This suggests that institutional arrangements and cultural influences are of major importance and indicates that further explanation for the significant difference is required. It also reinforces the presumption that there is huge potential for efficiency gains in the sickness absence insurance market.

Figure 1.1: Differences in Annual Absence Days by OECD Country



Depending on a country's institutional system, employers, private insurance companies, or social security systems provide sick pay. In the case of employer-provided sick pay, companies must bear the burden of labor costs in addition to productivity losses caused by absences from the workplace.

Under Germany's generous sick pay system, employers are legally obliged to continue to pay employees their full wage for up to six weeks per sickness episode. Unlike in most other countries, no benefit cap is applied. Nevertheless, as Figure 1.1 demonstrates, Germany is positioned in the middle region of the country ranking and some cross-country comparisons even place Germany below the international average in terms of sickness absence rates (Bonato and Lusinyan, 2004). One explanation might lie in the anecdotal evidence that Germans have a strong work ethic. Other explanations may be a well-functioning monitoring system or high unemployment rates.

In 1996, the Kohl government decided to reduce the statutory sick pay level from 100 to 80 percent of foregone gross wages, effective from October 1, 1996. The intention was twofold: to reduce moral hazard in the sickness absence insurance and to reduce labor costs in order to foster employment creation. At that time, employers had sick leave payments amounting to €28.2 billion per year, representing 1.5 percent of 1996 GDP (German Federal Statistical Office, 1998). Germany was positioned at the top of cross-country rankings comparing total labor costs per hour. At the time there was a consensus among economists that the extraordinarily high labor costs were the main reason for the persistently high unemployment rate in Germany.

The main aim of this chapter is to estimate the causal impact of the cut in statutory sick pay on sickness absence and labor costs. I exploit the exogenous variation in the absence costs by using a difference-in-differences methodology and longitudinal survey data from the German Socio-Economic Panel Study (SOEP). In the first part of the empirical analysis, I apply an intention-to-treat approach that compares those who were totally unaffected by the law—public sector employees, self-employed, and apprentices—to the target population of the reform: private sector employees. I thereby estimate the actual overall reform effect rather than the potential effect had the reform been strictly applied by every single company. The latter is not the case since a) employers are always free to provide fringe benefits on top of statutory regulations and b) persistent mass demonstrations and strikes forced employers' representatives in some industries to agree to the continuation of the old sick pay scheme in collective wage agreements. Thus, in the second part of the empirical analysis, I exploit differences in sick leave schemes across collective agreements of the most important industries. My data comprise individual-level information on the employees' industries and on whether individuals were covered by collective agreements. In the second part, I solely focus on private sector employees, using those who were covered by collective agreements with 100 percent guaranteed sick pay as controls. Contrasting them with employees in industries where sick pay was unambiguously cut as well as with employees that were not covered by collective agreements provides additional evidence on the reform's impact.

This chapter makes a contribution to the literature on the topic in several ways: I estimate the causal effects of cuts in sick pay levels on sickness absence behavior using non-Swedish and uncensored data. To identify the causal reform effect I make use of three different approaches that rely on different subsamples that were unequally affected by the reform. By this means, the plausibility and robustness of the identified effects is verified. While I provide evidence on the overall reform effects in the first part of the analysis, we gain greater insight into the specific mechanisms of the reform in the second part. Thanks to the panel structure of the data, I am able to take the sample composition into account. Most of the evaluation literature struggles with selection issues that often significantly hamper the analysis. In this context, I can identify job changers and employees who potentially selected themselves into or out of the treatment. Thus, sorting is unlikely to be a major issue. Unlike studies that estimate effects in certain regions or states, I use a representative sample of the most populous European country.

This chapter also contributes to the broader field of literature on the interdependencies between social insurance systems and labor supply. Since reduced labor costs was one of the main objectives of the reform, I calculate employers' total labor cost savings and roughly estimate the number of jobs which may have been created as a consequence of the reform.

Finally, this chapter illustrates the pitfalls that policymakers face when planning to implement unpopular labor market reforms in countries with a strong tradition of collective bargaining and strong unions. Had the purpose of the reform been better communicated and had the new law been applied one-to-one by all employers, my calculations suggest that twice as many jobs could have been created as actually occurred.

Section 1.2 outlines the institutional setting in Germany. Section 1.3 provides more detail on the data. Section 1.4 discusses the empirical estimation strategy. This is followed by Section 1.5 in which I provide the results of my empirical analyses. Section 1.6 outlines the chapter's conclusions.

1.2 The German Sick Pay Scheme and the Policy Reform

1.2.1 The Sick Pay Scheme and Monitoring System

Germany has one of the most generous sick pay schemes in the world. Before the implementation of the new law, every employer was legally obliged to continue usual wage payments for up to six weeks per sickness episode. In other words, employers had to provide 100 percent sick pay from the first day of a period of sickness with no benefit caps.[1] Henceforth, I use the term short-term sick pay as a synonym for employer-provided sick pay and short-term absenteeism as a synonym for periods of absence of less than six weeks.

In the case of illness, employees are obliged to immediately inform their employer about both the sickness and expected duration. From the fourth day of a sickness episode, a doctor's certificate is required and is usually issued for up to one week, depending on the illness. If the sickness lasts more than six continuous weeks, the doctor needs to issue a different certificate. From the seventh week onwards, sick pay is disbursed by the sickness fund and lowered to 80 percent of foregone gross wages for those who are insured under Statutory Health Insurance (SHI).[2]

The monitoring system mainly consists of an institution called *Medical Service of the SHI* (Medizinischer Dienst der Krankenversicherung). One of the original objectives of the Medical Service is to monitor sickness absence. German social legislation codifies that the SHI is obliged to call for the Medical Service and a medical opinion to clarify any doubts about work absences. Such doubts may arise if the insured person is short-term absent with unusual frequency or is regularly sick on Mondays or Fridays. Similarly, if doctors certify sickness with unusual frequency, the SHI may ask for expert advice. The employer also has the right to call for the assistance of the Medical Service and expert advice. Expert advice is based on available medical documents, information about the workplace, and a statement which is requested from the patient. If necessary, the Medical Service has the right to conduct a physical examination of the patient and to cut benefits.[3] In 2007, about 2,000 full-time equivalent and independent doctors worked for the medical service and examined 1.7 million cases of absenteeism (Medizinischer Dienst der Krankenversicherung (MDK), 2008).

[1] The entitlement is codified in the *Gesetz über die Zahlung des Arbeitsentgelts an Feiertagen und im Krankheitsfall (Entgeltfortzahlungsgesetz)*, article 3, 4.

[2] In principle, there is no limit on the frequency of sick leave spells. However, if employees fall sick again due to the same illness after an episode of six weeks, the law explicitly states that they are only again eligible for employer-provided sick pay if at least six months have been passed between the two spells or twelve month have been passed since the beginning of the first spell. This paragraph/clause intends to avoid substitution of long-term spells by short-term spells.

[3] The wording of the laws can be found in the Social Code Book V, article 275, para. 1, 1a; article 276

1.2.2 The Policy Reform

In 1996, the total sum of employer-provided sick pay amounted to DM 55.3 billion (€28.2 billion) (German Federal Statistical Office, 1998) and it was claimed to contribute to persistently high unemployment rates by functioning like a tax on labor. Together with speculation about a high degree of moral hazard in the generous German sick pay scheme, these considerations prompted the German government to pass Employment Promotion Act[4] which went into effect October 1, 1996.

The law reduced the sick pay employees are entitled to claim from 100 to 80 percent of gross wages during the first six weeks per sickness episode. The Self-employed were unaffected by the new law. Due to political considerations and the existence of other laws, both public sector employees and apprentices were exempt from the reform.[5] Similarly unaffected were employees on sick leave due to work related accidents. As an alternative to the cut in sick pay, from the date when the new law became effective, employees had the right to reduce their paid vacation by one day for every five days of sickness related absence, thereby avoiding the sick pay cut.

In addition to this law, which lowered short-term sick pay and is the focus of this chapter, another law was passed on November 1, 1996, and became effective from January 1, 1997, onwards.[6] The second law reduced sick pay from the seventh week onwards from 80 to 70 percent of forgone gross wages. The impact of this law on long-term absenteeism is analyzed in Chapter 3 of this thesis. So as to not confuse the impact of the cut in long-term sick pay with the impact of the cut in short-term sick pay, it is important to analyze the effects of both reforms separately. This is also necessary since the subgroups affected differed between the two reforms.[7]

Before, and in the aftermath of the law's implementation, the general public and the unions put pressure on employers through mass demonstrations and strikes. Germany is the country of origin of Bismarckian corporatism that has been adopted

[4] The *Arbeitsrechtliches Gesetz zur Förderung von Wachstum und Beschäftigung (Arbeitsrechtliches Beschäftigungsförderungsgesetz), BGBl. I 1996 p. 1476-1479*, was passed September 25, 1996.

[5] In the case of apprentices, the *Berufsbildungsgesetz (BBiG)* prevented the application of the law.

[6] This law is the *Gesetz zur Entlastung der Beiträge in der gesetzlichen Krankenversicherung (Beitragsentlastungsgesetz - BeitrEntlG), BGBl. I 1996 p. 1631-1633.*

[7] As an example, Puhani and Sonderhof (2010) do not differentiate between both reforms. In addition, they solely contrast group A.1 with group A.2 in Figure 1.2. In doing so they compare two groups that both include treated and non-treated employees.

as a role model by several European countries. An integral part of Bismarckian corporatism is the idea of social partnership between employers and unions along with autonomy in bargaining. As a result, unions are traditionally strong in Germany, as is the degree of collective bargaining coverage. In 1998, about 68 percent of all employees in West and 50 percent in East Germany were covered by collective wage agreements (Hans Böckler Stiftung, 2010). Ongoing union pressure forced employers' associations in various industries to agree in collective agreements to provide sick pay voluntarily on top of the statutory regulations. However, these collective agreements were only binding for firms under collective bargaining coverage.[8] Figure 1.2 gives an overview of how the reform worked.

Figure 1.2: Overview of Treatment and Control Groups

<table>
<tr><th colspan="3">Treatment intended by law (A)</th><th>No treatment intended by law (B)</th></tr>
<tr><td colspan="3">Private Sector Employees
(without apprentices)
(~25 milllion in 1995)</td><td>Apprentices, self-employed & public sector employees
(~1.5 + 3.5 + 5.5 million in 1995)</td></tr>
<tr><td>No collective agreement (A.1)
(~10 million)</td><td colspan="2">Collective Agreement (A.2)
(~15 million)</td><td rowspan="2"></td></tr>
<tr><td>Sick pay cut implemented at firm level; employer may provide fringe benefits on top of statutory regulations</td><td>Sick pay cut implemented at industry level by collective agreement (A.2.1)</td><td>Sick pay cut not implemented; employers in certain industries agreed to provide 100% sick pay (A.2.2)</td></tr>
<tr><td>Cut in sick pay</td><td>Cut in sick pay</td><td>No cut in sick pay</td><td>No cut in sick pay</td></tr>
</table>

Source: German Federal Statistical Office (1996, 2008); Hans Böckler Stiftung (2009); author's illustration

Principally, one needs to differentiate between those occupational groups that were intended to be affected by the weakening of statutory minimum standards and those that were not. The law applied to all employees in the private sector (A), whereas apprentices, self-employed, and public sector employees (B) were exempt as explained above. At the lower level, one needs to differentiate between those private sector employees who were covered by collective agreements (A.2) and those who were not (A.1). For companies not covered by collective bargaining, the decision on whether fringe benefits are provided on top of statutory standards is made at the company level. However, companies that are not covered by collective agreements are usually unlikely to provide a substantial amount of fringe benefits. Thus, it is very likely that those ten million private sector employees who were not covered by a collective

[8] Employers are free to leave collective agreements after the expiration of the contract. Average contract terms were 16.8 months in 1997; they even increased steadily from 1994 to 1998 (Hans Böckler Stiftung, 2010). In addition, I have not found any evidence that an unusual number of firms left collective agreements after the reform.

agreement experienced the cut in statutory sick pay. Those who were covered by collective bargaining were either not affected or were affected by the cut in sick pay, depending on the industry and the result of collective bargaining. In my data, I am able to identify all four groups as displayed in Figure 1.2. As shown below, I estimate three different empirical models that compare groups a.) A and B, b.) A.2.1 and A.2.2, and c.) A.1 and A.2.2 to identify the reform effects. I call these three models Approach I, Approach II, and Approach III.

1.3 Data And Variable Definitions

The empirical specifications make use of the German Socio-Economic Panel Study (SOEP). The SOEP is the only available representative data set for Germany that includes information on sickness absence. The SOEP is a longitudinal annual household survey that has existed since 1984. Wagner et al. (2007) provide further information.

For the empirical specifications, I extract three pre- and two post-reform years from the survey, i.e., the waves from K (1994) to P (1999) that each contains sickness absence information about the previous year.[9] I discard the year 1996 in most specifications because the reform went into effect October 1, 1996, and because I only have absence information based on calendar years. However, in one robustness check, I estimate treatment effects separately for the years 1996, 1997, and 1998.

I restrict my sample to those in the labor force who are eligible for sick pay (plus self-employed) and who are between 18 and 65 years of age.[10] Work accident related sick leave was exempted from the new regulations. Thus, I exclude all respondents who needed medical treatment due to a work accident in one of the sampling years. Besides short-term sick pay, long-term sick pay, which is disbursed from the seventh week onwards, was also effectively reduced as of January 1997. Since I intend to isolate the reform effects on short-term absenteeism, I discard all respondents hav-

[9] What is meant here is that I collect data from the years 1993-1998. Since current as well as retrospective information is sampled in every wave, I match the retrospective information which I am interested in with the current information of the relevant year as long as the respondent was interviewed in both years. If this was not the case, I use both types of information from the same interview and assume that the current statements have not changed since the previous year.

[10] Although marginally employed (employees who earn less than €400 per month) are eligible for sick pay and have been on a par with the full-time employed since June 1, 1994, I do not include them since it is likely that marginally employed were not fully aware of their rights at that time and since anecdotal evidence suggests that a significant proportion of employers refused to provide this benefit.

ing had a long-term sickness spell of more than six weeks in one of the sampling years.[11] In the empirical assessment, I provide evidence suggesting that excluding the homogeneous, but special, group of the long-term sick poses no severe threat to the evaluation of the cut in short-term sick pay. I also discard individuals with item non-response.

1.3.1 Endogenous and Exogenous Variables

The SOEP is a rich data set, particularly with respect to job characteristics. Detailed questions about the type of job, the number of years with the employer, the gross and net wage, etc. are sampled. Additionally, there are questions on sick leave behavior.

I generate my dependent variables from the following question: *"How many days off work did you have in 19XX because of illness? Please enter all days, not just those for which you had a doctor's certificate." a.) No days b.) XX days in total.* The great advantage of the SOEP and this question is that the *total* number of absent days is documented, not only those with a certificate or those that are compensated by a federal institution as it is the case with most register data. Particularly when the focus is on short-term absenteeism, it is a big advantage to have such a total measure. However, this comes at the cost of not having detailed spell data.

The first dependent variable measures the proportion of employees with no absence days, i.e., the incidence of sickness absence. It is named *noabs* and has a one for employees with no sick leave days; all employees with sick leave are coded as zeros. *Noabs* should not be very prone to measurement errors.

The second dependent variable measures the total number of absent days and is called *daysabs*. However, looking at the distribution of this variable, the potential issues of measurement errors, misreporting behavior, and outliers become quite obvious. For example, 0.2 percent (i.e., 30 respondents) of the sample indicated a total number of absence days of more than 50, which is theoretically possible but, given that these respondents also denied an absence spell of more than six weeks, quite unlikely. While the evaluation of the reform effects should not be seriously distorted as long as the reform did not affect measurement errors, outliers and misreporting

[11] The identification of these respondents is feasible since a question on whether respondents had such a long-term spell is annually asked. In Section 1.5.1, I again use the whole sample to estimate the total labor cost savings for Germany. Likewise, respondents are asked every year whether they had a work accident and whether this accident required medical treatment.

potentially exacerbate standard errors and lead to imprecise estimates. Moreover, in the case of underreporting, with $0 < \alpha < 1$ being the reported fraction of true days, my estimates would be downward biased by α.

The entire set of explanatory variables can be found in Appendix A and are categorized as follows: the first group incorporates variables on personal characteristics, such as the dummies *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 higher education degree awarded, number of years in current workplace, and whether the person was trained specifically for their job. The last group contains explanatory variables on job characteristics. Among them are *blue-collar worker, white-collar worker*, the size of company, or *gross wage per month*. I also control for the annual state unemployment rate and include state as well as year dummies.

1.3.2 Treatment and Control Groups

As already mentioned in Section 1.2.2 and visualized in Figure 1.2, my main approach contrasts private sector employees (A), which were intended to be treated by the reform, with all unaffected occupational groups (B). I call this Approach I ("intention-to-treat approach") and generate a dummy variable *T1* that has a one for all respondents in group A of Figure 1.2 and a zero for all respondents in group B.

Besides the main specification, I apply two more specific models in order to better pinpoint the exact mechanisms of the reform. In these models, I solely focus on private sector employees. In Approach II, I compare private sector employees whose collective agreements codified a cut in sick pay from 100 to 80 percent for the first three days of a sickness spell (A.2.1) to private sector employees whose collective agreements codified 100 percent sick pay (A.2.2). In Approach III, I compare private sector employees who were not covered by collective agreements (A.1) to group A.2.2 (see Figure 1.2).

Performing Approach II and Approach III is only feasible since I have the following information: after an extensive review of all collective agreements in the main industries, I identified seven main industries that completely excluded any changes in the sick leave regulations, e.g., the chemical industry or credit and insurance industries (A.2.2). Likewise, I was able to identify industries that unambiguously implemented a cut in sick pay from 100 to 80 percent for the first three days of a sickness spell, e.g., the construction and agriculture sector (A.2.1).[12]

Since SOEP data provide 2-digit industry codes, I am able to identify employees working in these industries. Moreover, the SOEP includes information on whether employees were covered by collective bargaining. In 1995, the year prior to the reform, this information was sampled. Hence, for Approach II and Approach III, I use respondents who answered the collective bargaining question in 1995 and keep them in for all years in which they held the same job as in 1995. In this way, I exclude at the same time job switchers who might have selected themselves out of the treatment and hence control for treatment-related selection.

For Approach II, I generate a dummy, *T2,* that is one for all employees who were covered by collective agreements and worked in industries that strictly implemented a cut from 100 to 80 percent for the first three days of a sickness episode (A.2.1). *T2* is zero for all other respondents in the subsample, i.e., private sector employees who did not experience changes in their sick pay schemes (A.2.2).

For Approach III, I generate a dummy named *T3* that is one for all private sector employees who were not covered by collective agreements (A.1). *T3* is zero for all private sector employees who were covered by collective agreements and worked in industries that excluded any changes in the sick leave schemes in their agreements (A.2.2).

[12] In many industries unions successfully prevented a cut in replacement levels since this had a strong symbolic character for most unions. However, in return for that, in these industries other changes in the sick leave schemes were implemented such that benefits were effectively cut as well. For example, a popular consensus was to keep the replacement rate at 100 percent but to exclude overtime hours from the basis of calculation (Hans Böckler Stiftung, 2010). Before the reform, overtime hours were usually considered in calculating the average gross wage to which the replacement rate was applied. Since my approaches II and III intend to unfold precise reform mechanisms, I decided to only use industries where cuts were unambiguously enforced or in which sick leave schemes were not changed at all. In that respect, Figure 1.2 provides a simplified illustration. To be precise, in many industries in A.2.2 replacement levels were not cut but sick leave benefits were decreased indirectly. In 1998, union leaders proudly declared that 13 million employees would receive 100 percent sick pay (Jahn, 1998)—a statement that concealed that, nevertheless, many of them effectively experienced cuts in their sick pay.

1.4 Estimation Strategy and Identification

1.4.1 OLS Difference-in-Differences Model

I start by estimating conventional difference-in-differences (DiD) models of the following form by OLS:

$$y_{it} = \lambda post97_t + \pi D_{it} + \theta DiD_{it} + s'_{it}\psi + \rho_t + \phi_s + \epsilon_{it} \tag{1.1}$$

where y_{it} stands either for the incidence of sickness absence (*noabs*) or for the annual number of absence days of individual i in year t (*daysabs*). The variable, $post97_t$, is a post-reform dummy, D_{it} is a treatment indicator (*T1*, *T2*, or *T3*), and DiD_{it} is the regressor of interest. It is one for treated respondents in post-reform years and gives us the causal reform effect if certain assumptions hold. It can also be interpreted as the interaction term between the treatment indicator and the post-reform dummy. By including additional time dummies, ρ_t, I 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 is included in the $K \times 1$ column vector s'_{it}. The other $K-1$ regressors are made up of personal controls including health status, educational controls, and job-related controls as shown in Appendix A. As usual, ϵ_{it} stands for unobserved heterogeneity.

1.4.2 Identification

In each model, I compare individuals who were affected by the reform to individuals who were not affected by the reform. In doing so I analyze changes in the outcome variables over time for the treatment group relative to the control group. At the same time, the sample composition is adjusted for differences in covariates. This is the basic setup of all DiD analyses.

Then the main identification assumption states that all relative changes of the outcome variable of the treatment group depend entirely on the exposure to the reform. In other words, conditional on the available covariates, I assume the absence of unobservables with a differential impact on the work absence *dynamic* for treatment and control group. This is also called the common time trend assumption.

Table 1.1 shows in columns (1) and (2) the covariates' means separately for individuals with *T1=1* and *T1=0*, i.e., for samples A and B in Figure 1.2. It is easy to see that both groups differ with respect for most of the observables. Imbens and Wooldridge (2009) have shown that if treated and control units differ substantially in their observed characteristics, then parametric approaches use the covariate distribution of the controls to make out-of-sample predictions that might lead to sensitive results. Moreover, more homogenous samples yield more precise estimates. Combining matching and regression outperforms pure matching or regression approaches and yields more robust results (Imbens and Wooldridge, 2009).

Table 1.1: Sample Means and Normalized Differences of Raw and Matched Sample

	Raw Sample			**Matched Sample**		
Covariates	**Treated mean**	**Controls mean**	**Norm. diff.**	**Treated mean**	**Controls mean**	**Norm. diff.**
Outcome variables						
Noabs	0.524	0.527	0.021	0.528	0.533	0.008
Daysabs	5.105	5.268	0.008	5.122	5.259	0.007
Personal covariates						
Female	0.400	0.490	0.121	0.416	0.463	0.068
Age	39.29	38.71	0.108	39.24	38.88	0.024
Immigrant	0.205	0.096	0.213	0.184	0.108	0.148
East	0.231	0.309	0.119	0.250	0.311	0.098
Partner	0.799	0.705	0.219	0.789	0.736	0.088
Married	0.697	0.629	0.165	0.688	0.651	0.059
Children	0.478	0.468	0.008	0.474	0.478	0.005
Disabled	0.029	0.032	0.000	0.031	0.033	0.006
Health good	0.678	0.655	0.011	0.675	0.650	0.036
Health bad	0.065	0.070	0.011	0.067	0.069	0.007
No sports	0.409	0.337	0.123	0.408	0.348	0.093
Educational covariates						
Dropout	0.044	0.031	0.006	0.043	0.033	0.032
8 years of schooling	0.332	0.234	0.169	0.317	0.264	0.085
10 years of schooling	0.331	0.390	0.066	0.346	0.400	0.080
12 years of schooling	0.043	0.045	0.008	0.039	0.048	0.032
13 years of schooling	0.144	0.254	0.197	0.156	0.205	0.088
Other certificate	0.107	0.045	0.170	0.100	0.051	0.131
Trained for job	0.563	0.573	0.000	0.564	0.572	0.007
Job covariates						
Part-time employed	0.144	0.132	0.041	0.144	0.134	0.017
New job	0.138	0.112	0.037	0.135	0.116	0.035
Years with company	8.828	9.415	0.044	8.879	9.248	0.028

Continued on next page...

... *Table 1.1 continued*

Covariates	**Treated mean**	**Controls mean**	**Norm. diff.**	**Treated mean**	**Controls mean**	**Norm. diff.**
Small company	0.277	0.245	0.051	0.276	0.243	0.050
Medium company	0.302	0.216	0.135	0.302	0.216	0.141
Large company	0.227	0.205	0.036	0.229	0.213	0.027
Very large company	0.194	0.259	0.106	0.192	0.254	0.103
White collar	0.564	0.381	0.305	0.540	0.429	0.164
High job autonomy	0.199	0.342	0.227	0.194	0.317	0.200
Gross wage per month	2,031	2,029	0.043	2,016	1,980	0.024
State unemployment rate	11.262	12.001	0.134	11.423	11.951	0.095

"Treated" stands for *T1*=1 and "Controls" stands for *T1*=0, i.e., samples A and B in Figure 2 are compared. "Norm. Diff." stands for "normalized difference" which is calculated according to $\frac{\bar{\mu}_1-\bar{\mu}_0}{\sqrt{\sigma_1^2+\sigma_0^2}}$, where $\bar{\mu}_1$ is the sample mean of the covariate for the treatment group and $\bar{\sigma}_0^2$ stands for the variance of the covariate within the control group. The "matched sample" is generated by means of five-to-one nearest neighbors matching based on the propensity score. The propensitiy score (PS) is the probability of belonging to the treatment group and is estimated by a logit model with the inclusion of the following covariates: *female, immigrant, partner, married, disabled, 9 (10, 12, 13) years of schooling, trained for job, new job, years with company, white collar, gross wage per month, state unemployment rate*. To estimate the propensity score, covariates are selected according to likelihood ratio tests on zero coefficients. After the PS estimation, in the matched sample, 2,897 observations are not included since they are not assigned a nearest neighbor due to extreme PS-values. In total, the raw sample contains 20,700 observations and the matched sample contains 17,803 observations.

Imbens and Wooldridge (2009) propose to judge the differences in covariates for treatment and control group by the scale-free normalized difference $\Delta\mu = \frac{\bar{\mu}_1-\bar{\mu}_0}{\sqrt{\sigma_1^2+\sigma_0^2}}$ with σ_1^2 being the variance for the treated and $\bar{\mu}_0$ being the mean for the controls. As a rule of thumb, a normalized difference exceeding 0.25 is likely to lead to sensitive results. Column (3) shows that for 15 covariates, the normalized difference between the main treatment and the main control group is larger than 0.1 and that for 4 covariates, it is even greater than 0.2. To achieve a better balance across covariates, I apply propensity score matching to the raw sample. More precisely, I perform five-to-one nearest neighbor matching on the probability of belonging to group A versus group B in Figure 1.2. Matching covariates were selected out of the whole set of covariates in Appendix A according to likelihood ratio tests on zero coefficients (see notes to Table 1.1 for more details). This way I obtain the "matched sample" as shown in columns (4) to (6). The matched sample only includes individuals who are similar to one another in terms of their observable characteristics. It contains 2,897 observations fewer than the raw sample and has much better balancing properties.

Except for one case, all normalized differences are below 0.2. I use the matched sample to estimate my main model, which compares individuals in group A to similar individuals in group B (Approach I).

In addition to applying propensity score matching in the first step to make the treatment and control groups in my intention-to-treat approach as comparable as possible, I have additional arguments why the common time trend assumption is very likely to hold in this case: first, I incorporate a rich set of covariates in the regression models that controls for differences in personal, educational, and job characteristics. It should be emphasized that I observe the (self-reported) health status, sporting activities, and disability status of the respondents. Correlating absence days and all available control variables shows that age, good health, schooling, and high job autonomy are negatively correlated with sick leave. In line with the literature, males and part-time employees have fewer absence days while bad health and company size is positively correlated with absenteeism. High regional unemployment rates serve as a worker discipline device in the sense of Shapiro and Stiglitz (1974). All factors that the empirical literature has identified as important determinants of sickness behavior can be controlled for. In the DiD models, I also take time-invariant sick leave differences of the treatment and control group into account and adjust for time trends as well as state-specific effects.

Second, in the results section, I present the results of placebo regressions. Placebo regressions assume that the reform under analysis took place in a year without reform. Should the coefficient of interest be significant in a non-reform year, the common time trend assumption would be seriously challenged.

Finally, I plot the development of the incidence and duration of sick leave for treated and controls over time. In my basic intention-to-treat model where I compare private sector employees with other unaffected occupational groups, critics could argue that the common time trend assumption would be questionable since the different occupational groups would face different economic incentives to attend work—despite having controlled for a rich set of socioeconomic background characteristics, time and state trends as well as dynamics in the regional unemployment rates. Figures 1.3 and 1.4 show the evolution of the two main outcome variables for the matched sample in Table 1.1 from 1993 to 1998.[13]

[13] As explained above, I exclude 1996 from my analyses because the reform became effective from October 1, 1996 onwards and because I only measure the total number of calendar year days absent. Thus, in both figures, the year 1996 does not represent an observation. Unfortunately, workplace

Figure 1.3 shows *noabs* and thus represents the proportion of employees without absence days. We observe that the curves for treated and controls run clearly parallel for the three pre-reform years. From 1995 to 1997, we observe a distinct increase in the share of private sector employees without any absence days, while the share remains relatively stable for the controls. From 1997 to 1998, we again observe a parallel development of both curves.

Figure 1.3: Share of Non-Absent Respondents for Treatment and Control Group Over Time

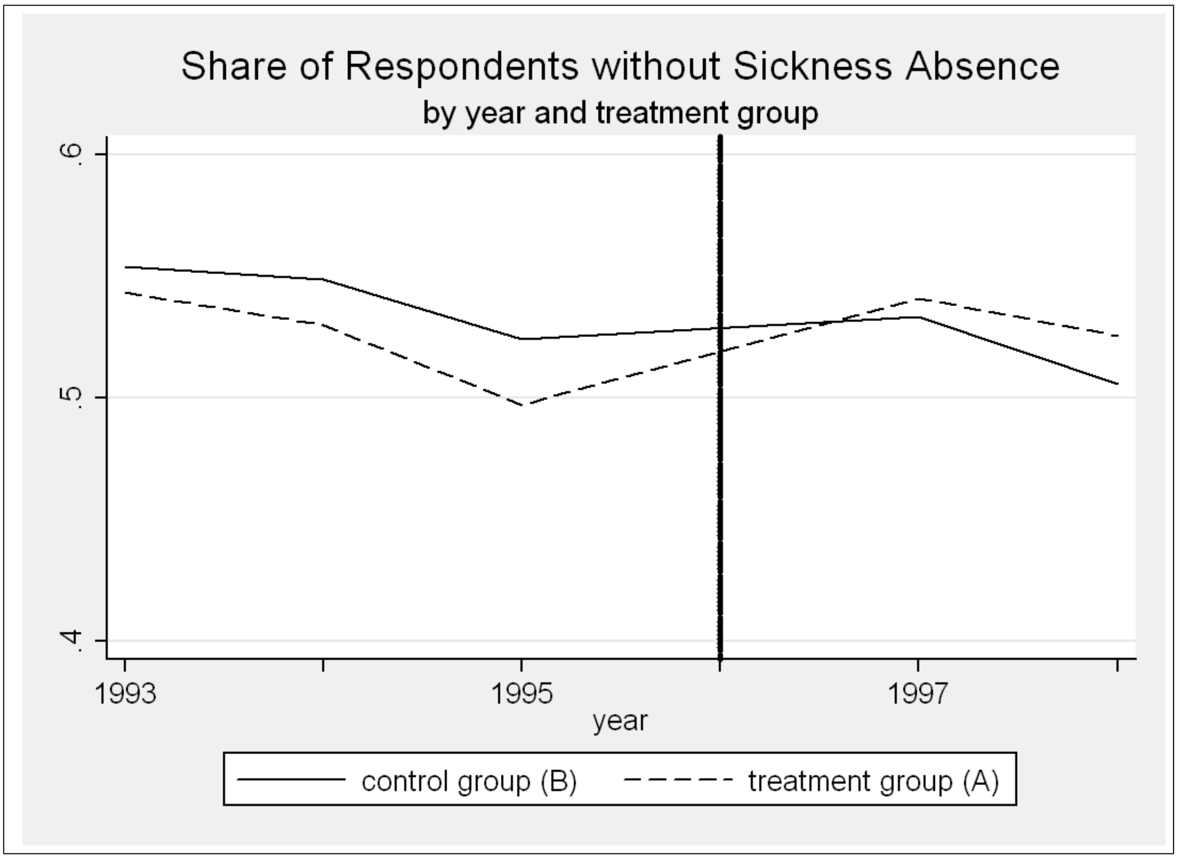


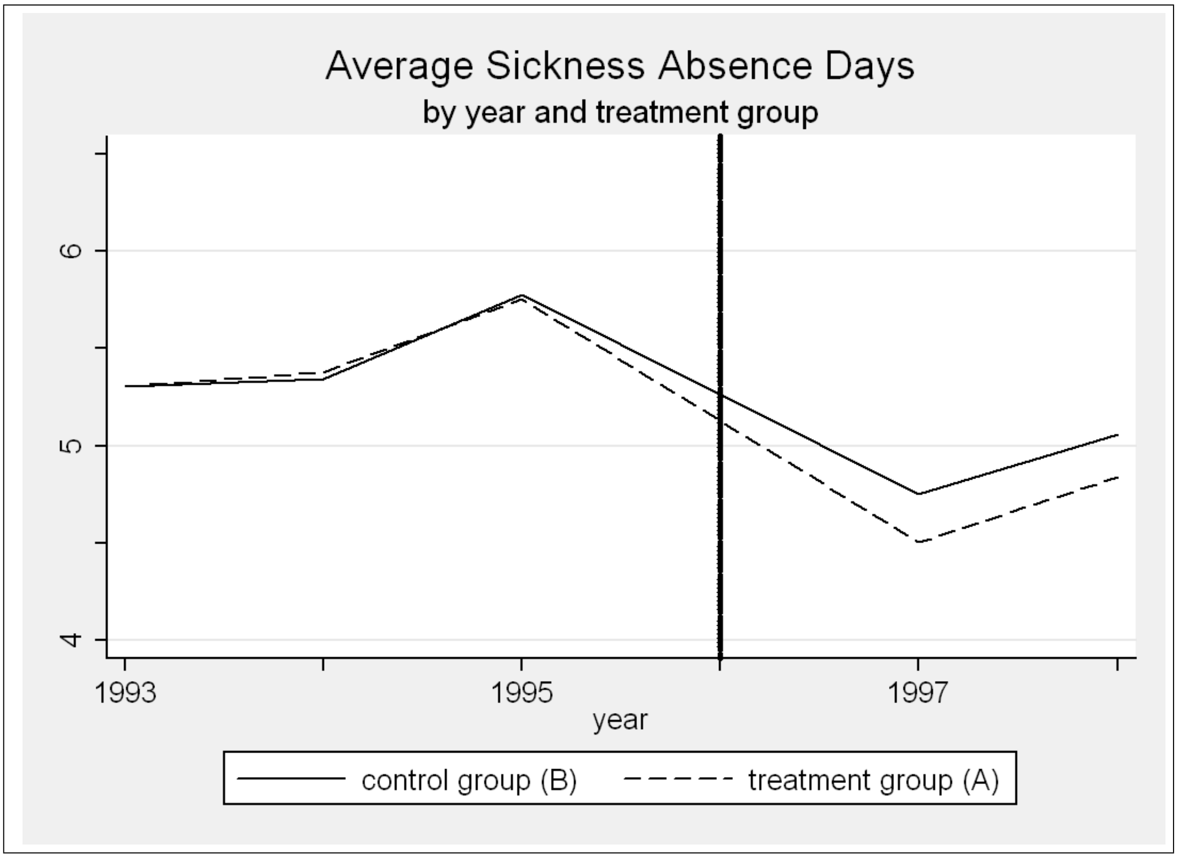
Figure 1.4 draws a very similar picture for the annual number of absence days. We see a remarkably parallel development of *daysabs* for both groups and the pre-reform years. From 1995 to 1997, both curves decline but the decrease for the treated was much more pronounced. From 1997 to 1998, we again observe parallel curves.

Both figures strongly support the assumption of common time trends. Moreover,

absences was not sampled in 1992 and data after 1998 cannot be used because laws were changed effective January 1, 1999.

it should be kept in mind that both figures draw raw, unconditional pictures. In the parametric regressions, I additionally correct for various influence factors as detailed in equation (1.1).

Figure 1.4: Average Sick Leave Days for Treatment and Control Group Over Time



There is an intensive debate about the drawbacks and limitations of DiD estimation. One particular concern is the underestimation of OLS standard errors due to serial correlation in the case of long time horizons and unobserved (treatment and control) group effects. As Bertrand et al. (2004) show, the main reason for the understating of standard errors is rooted in serial correlation of the outcome and the intervention variable and is basically eliminated when focusing on fewer than five periods. While there is consensus about the serial correlation problem, the issue of unobserved common group effects is more controversial. If one takes the objection of Donald and Lang (2007) seriously, then it is not possible to draw inferences from DiD analyses, in the case of few groups, meaning that no empirical assessment could be performed.

Wooldridge (2007) asks rhetorically whether introducing more than sampling error into DiD analyses is necessary or desirable and whether we should conclude that nothing can be learned in such cases. Angrist and Pischke (2009) provide an excellent discussion on the various approaches of how to correct standard errors in general, and in DiD models in particular. Besides the Donald-Lang correction, clustering at a higher aggregated level is considered as one possibility; however, as a rule of thumb, at least 42 clusters are needed.

Alongside my focus on short time spans to resolve serial correlation concerns, I use robust standard errors and, in Approach I, I cluster observations on the individual level. In Approach II and III, I cluster on the industry×year level, since this is the level at which negotiations about the application of the reform took place. Moreover, for all approaches, I demonstrate how standard errors change when I apply the Donald and Lang (2007) two-step procedure to correct for unobserved group errors.

One of the biggest issues in evaluation studies is self-selection. Here, the reform was politically determined and the law applied to all private sector companies. It is very unlikely that people left the labor market due to the cut in sick pay. Selection out of the treatment in the sense that a substantial amount of Germans became self-employed (with no sick pay at all) or public sector employees is equally unlikely. However, information on whether people changed their jobs and information on the labor market status allows me to control for this possibility. As explained above, in the specific approaches II and III, I need to exclude all job changers and solely rely on employees who held to same job as in the pre-reform year 1995.

There may also be concerns about the policy change being endogenous in the sense that the reform was a reaction to increasing absence rates (Besley and Case, 2000). I find no evidence that this might have been the case. The reform was not a reaction to increasing absence rates but rather a tool for reducing the persistently high labor costs that were institutionalized. The reform was random insofar it was mainly an instrument used by the unpopular Kohl government (which had been in power since 1982) to demonstrate its strength and capacity to act. Structural reforms of the employer sick leave pay system had been debated in Germany since the beginning of the 1980s (Lambsdorff, 1982).

However, I cannot exclude the possibility that the actual reform enforcement by specific industries as discussed in Section 1.2.2 was endogenous and related to absence rates. For example, it is possible that unions in industries with a high structural number of absence days fought harder for fringe benefits and 100 percent sick pay. On the other hand, it is also possible that employers in such industries insisted on the cut in sick pay. In the first case, I would underestimate, and in the second case overestimate, the reform effects. In any case, I do not find any systematic relationship between industry absence rates and the embodiment of sick pay schemes in collective agreements (Hans Böckler Stiftung, 2010). Industry-specific unobservables are not serious threats to my estimates as long as they have no impact on sick leave dynamics over time. For example, there should not be post-reform changes in industry-specific sick leave monitoring that are correlated with negotiation outcomes. As long as all employers in a specific industry did not systematically alter sick leave rules, the enforcement of the reform is exogenous to the individual employee. To obtain estimates that are as clean as possible, I review all collective agreements industry-by-industry. Precisely because unions did make concessions in order to keep the main sick leave pay level at 100 percent, I include only those industries in group A.2.2 (see Figure 1.2) that definitely made no other concessions to the industry-specific sick pay schemes. Hence, in my Approach II, the decision to cut sick leave is made at the *industry*-level by few employee and employer representatives, while in Approach III, the decision to cut sick leave was made at the *company*-level. Hence, employers in both groups should reveal, if any, diverse behavioral reactions to the reform. These would translate into differences in the estimates. The fact that my empirical findings for Approach II and Approach III are very close suggests that post-reform changes in employers' behavior that are correlated with the level of sick pay are unlikely to play a substantial role.

As discussed, the evaluation of the reform rests upon three different approaches. Approach I is my main specification, which I called "intention-to-treat approach." It contrasts group A in Figure 1.2 with group B—after having made these groups comparable by means of propensity score matching. With this approach I intend to capture the overall reform effects for the whole of Germany. My two more specific approaches, II and III, intend to isolate some of the specific reform mechanisms. Approach II contrasts those who experienced a cut in sick pay from 100 to 80 percent in the first three days of a sick leave episode with totally unaffected private sector employees (A.2.1 vs. A.2.2). Approach III contrasts those who were not covered by collective agreements (A.1) with unaffected employees in the private sector (A.2.2). Although not all of these groups are mutually exclusive, the three approaches make use of three different treatment groups that were all differently affected by the reform. Likewise, two different control groups are used. The findings from these three approaches in combination with the different outcome measures and my robustness checks should yield enough material to evaluate the reform. Furthermore, the pattern of results in combination with knowledge about the German labor market and anecdotal evidence about the events that accompanied the reform should yield evidence on the robustness and plausibility of my identification strategy.

1.5 Results

1.5.1 Intention-to-Treat Approach (Approach I)

Part A of the empirical analysis assesses the overall reform effect among those for whom treatment was intended. Using the matched sample of Table 1.1 we compare the sick leave behavior of group A with the sick leave behavior of group B in Figure 1.2 over time. The first step of Part A is to run models that evaluate whether the incidence of sick leave changed as a result of the increase in absence costs.

Effect on Share of Non-Absent Employees

As already discussed in the previous section, Figure 1.3 plots the share of workers without absence days. The graph already provides support for the notion that the share of employees without any sick leave days increased due to the reform. Taking the raw 1993-1995 figures against the raw 1997-1998 figures for both groups—which are plotted in Figure 1.3—the unconditional difference-in-differences estimate is +3.43 percentage points and significant at the 2.7 percent level.

Table 1.2 shows the results for this exercise when sets of covariates are added piecewise from the first to the fourth column. All models yield highly significant DiD estimates that are very close to each other in magnitude. From the preferred specification in column (4) we infer that the reform increased the share of employees with zero sick leave days by 3.4 percentage points. This estimate is significant at the 2.6 percent level. Related to the average pre-reform share of employees in the treatment group without absence days, which was 52.45 percent, this figure translates into a reform-induced increase in the share of non-absent employees by 6.5 percent. Please note that this estimate is not sensitive to functional form assumptions. Running a count data specification yields a highly significant reform effect of 4.1 percentage points which represents an increase of non-absent private sector employees of 7.8 percent.[14]

Four main conclusions can be drawn from Table 1.2: first, all estimates are in line with our expectation of how the reform might have affected the incidence of sick leave. Using the estimates from the preferred specifications, we conclude that the reform led to an increase in the share of employees without any sick leave days by between 6 and 8 percent. Second, all estimates are significant at the two or three percent level. Third, the estimates are very robust to the inclusion of different sets of covariates and very similar to the unconditional DiD estimate. This adds additionally to the credibility of the assumption of common time trends. Finally, the results are very similar in size and all lie within the same confidence intervals, no matter whether I run OLS or count data models. This suggests that functional form assumptions do not seem to matter very much.

Effect on Sick Leave Duration

The second step of Part A is to evaluate the overall reform effects on the average number of sick leave days. Table 1.3 displays the results when *daysabs* is used as the dependent variable. The first two columns show results for OLS-DiD models, where column (1) abstains from covariates and column (2) uses the whole set of covariates. For both models, I find a decrease of about 0.2 absence days which is, however, imprecisely estimated. As discussed in Section 1.3.1, these imprecise estimates are very likely attributable to extreme outliers and measurement errors in the *daysabs* variable as the following exercise shows: when I censor the dependent variable artificially at 50, the standard error decreases to 0.24 and the p-value decreases to 0.15. Censoring at 30 lets the standard error shrink further and the estimate is then significant at the 9 percent level. When I drop those 265 observations (1.5 percent of the sample) that indicated an implausibly large number of sick leave days greater than 30, and thus likely to comprise a substantial amount of measurement errors, I obtain an average decrease of about -0.44 days that is significant at the 3.2 percent

[14] More precisely, I run a Zero-Inflated Negative Binominal-2 Model. When I run the simple OLS difference-in-differences model with the "raw" instead of the "matched" sample of Table 1.1, I obtain an estimate of 2.5 percentage points which is significant at the 7 percent level. When I run the same model as in column (4) of Table 1.2 but cluster at the state×year instead of the individual level, the standard error increases to 0.0187 and the estimate is significant at the 7.5 percent level.

Table 1.2: Intention-to-Treat Approach: DiD Estimation on the Share of Non-Absent Employees Using Matched Sample

	Dependent variable: ***noabs***			
Variable	**(1)**	**(2)**	**(3)**	**(4)**
DiD	**0.0342****	**0.0329****	**0.0322****	**0.0337****
	(0.0155)	**(0.0152)**	**(0.0152)**	**(0.0151)**
	[0.028]	**[0.030]**	**[0.034]**	**[0.026]**
Post reform dummy	-0.0542***	-0.0590***	-0.0582***	-0.0819***
	(0.0142)	(0.0141)	(0.0140)	(0.0207)
Year 1997	0.0210**	0.0201**	0.0202**	0.0182*
	(0.0095)	(0.0095)	(0.0095)	(0.0097)
Year 1995	-0.0396***	-0.0387***	-0.0377**	-0.0454***
	(0.0108)	(0.0108)	(0.0108)	(0.0112)
Year 1994	-0.0117	-0.0125	-0.0111	-0.0204*
	(0.0102)	(0.0101)	(0.0101)	(0.0111)
Treatment Group	-0.0191	-0.0255**	-0.0283**	-0.0159
	(0.0121)	(0.0117)	(0.0117)	(0.0118)
Job characteristics	no	no	no	yes
Educational characteristics	no	no	yes	yes
Personal characteristics	no	yes	yes	yes
Regional unemployment rate	no	yes	yes	yes
State dummies	no	yes	yes	yes
Mean treated	0.5245	0.5245	0.5245	0.5245
Change in %	+6.5	+6.3	+6.1	+6.5
N	17,803	17,803	17,803	17,803
R^2	0.0012	0.0501	0.0523	0.0755

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on personal identifiers; p-values are in square brackets. All models make use of the matched sample in Table 1.1 and compare those who were intended to be treated—private sector employees—with those who were entirely unaffected by the reform (Approach I; group A vs. B in Figure 2). Every column stands for one OLS-DiD regression model as in equation (1.1). The dependent variable in all models is *noabs* (see section 1.3.1).

Table 1.3: Intention-to-Treat Approach: DiD Estimation on the Number of Absence Days Using Matched Sample

	Dependent Variable: *daysabs*							
			Quantile regression					
	OLS	**OLS**	**q=0.5**	**q=0.6**	**q=0.7**	**q=0.8**	**q=0.9**	**q=0.99**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
DiD	-0.2146	-0.2037	-0.1586*	-0.4015**	-0.5983*	-0.6811	-0.5596	0.6313
	(0.2872)	(0.2718)	(0.0944)	(0.2079)	(0.3303)	(0.4689)	(0.7783)	(3.5491)
Job characteristics	no	yes	yes	yes	yes	yes	yes	yes
Educational characteristics	no	yes	yes	yes	yes	yes	yes	yes
Personal characteristics	no	yes	yes	yes	yes	yes	yes	yes
Regional unemployment rate	no	yes	yes	yes	yes	yes	yes	yes
Time dummies	no	yes	yes	yes	yes	yes	yes	yes
State dummies	no	yes	yes	yes	yes	yes	yes	yes
Mean treated	5.45	5.45	1.27	3.13	5.49	8.72	14.35	31.66
Change in %	-3.9	-3.7	-12.5	-12.8	-10.9	-7.8	-3.9	+1.9
N	17,803	17,803	17,803	17,803	17,803	17,803	17,803	17,803
(Pseudo)R^2	0.0013	0.0635	0.0241	0.0594	0.0737	0.0752	0.0859	0.0875

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on personal identifiers in columns (1) and (2). All models make use of the matched sample in Table 1.1 and compare those who were intended to be treated—private sector employees—with those who were entirely unaffected by the reform (Approach I; group A vs. B in Figure 1.2). Columns (1) and (2) estimate OLS-DiD models as in equation (1.1). Columns (3) to (8) estimate DiD quantile regressions; marginal effects for these models are calculated at the means of the covariates except for *post97* (=1), *T1* (=1), *Year 1994* (=0), *Year 1995* (=0), *Year 1997* (=1), and *DiD* (=1). Every column stands for one regression model. The dependent variable in all columns is *daysabs* (see section 1.3.1).

level.

Another alternative to circumvent issues due to large standard errors is to run quantile regressions. At the same time, quantile regressions allow us to evaluate how different parts of the sick leave days distribution might have reacted with regard to the cuts in sick pay. Columns (3) to (8) of Table 1.3 display the results. I only find significant effects for the lower tail of the sickness day distribution. When I relate the decrease in absence days at the various quantiles to the mean number of absence days at these quantiles, a clear pattern emerges. The reform effect distinctly shrinks in size and significance the more one moves to the upper tail of the sickness day distribution. For employees with up to 5.5 average absence days per year (i.e., for q=0.5, q=0.6, and q=0.7), the effect is statistically significant and is about -12 percent. For q=0.8 onwards, the effects become insignificant and also decrease in relative magnitude. For the 99^{th} quantile, the estimate is even slightly positive.

All evidence taken together suggests that the bulk of behavioral reactions towards the cut in sick pay was triggered by employees in the lower tail of the sick leave days distribution. This insight is supplemented by Table 1.2 and the finding that the share of employees with zero absence days increased by between six and eight percent. Employees with up to 5.5 absence days per year decreased their use of sick leave by about 12 percent as a result of the cut in sick pay.

Robustness Checks and Effect Heterogeneity

In the final step of Part A I present robustness checks and results on effect heterogeneity as well as on placebo regressions. Panel A of Table 1.4 shows robustness checks using *noabs* as dependent variable. Column (1) excludes all respondents who indicated that they changed their jobs in the calendar year prior to the interview. By doing this I intend to shut down the possibility that selection out of the treatment in form of job switching might impact my results. Column (2) weights the regressions with the inverse probability of not dropping out of my sample in the post-reform period and abstains from a refreshment sample which was drawn in 1998. In doing so, I intend to control for labor market and panel attrition as well as panel composition effects. Both estimates are significantly different from zero. The size of the effects is similar to the reference estimate in column (4) of Table 1.2, although the estimate in column (2) of Table 1.4 is a little bit larger.

The next two columns of Panel A add 1996 to the analysis. In all other models, I omit this year because private sector employees were semi-treated as the reform went into effect October 1 that year. Column (4) estimates reform effects separately for each year. Estimating yearly effects decreases the precision of the estimates slightly. However, all three estimates are close to each other and lie between 2 and 3.3 percentage points. The fact that the average treatment effect for 1996 is about the same size as the effect in 1998 is not really surprising since the awareness about the reform among employees was probably greatest in 1996. Moreover, in 1996, uncertainty about the law's enforcement was also highest since it took some time before unions and employer representatives reached agreements in collective bargaining.

As discussed above, I excluded all respondents from my sample who had at least one sickness spell greater than six continuous weeks in the years under consideration. In this way we avoid confusing the effects of the cut in short-term sick pay (up to six weeks per spell) with the effects of a cut in long-term sick pay (from the seventh week onwards) that was enacted at the same time. However, this approach might be problematic should there be an effect on long-term absenteeism. Thus, the last column of Panel A makes use of the full sample, i.e., includes all respondents with long-term spells.[15] Then I estimate the reform effect on long-term absenteeism using the same model as in equation (1.1) with the incidence of long-term absenteeism as dependent variable. The effect is close to zero in magnitude and not significantly different from zero.

Figure 1.5 provides additional graphical evidence suggesting that omitting the fraction of respondents who are prone to long-term absenteeism does not seriously hamper the evaluation of the cut in short-term sick pay. Figure 1.5 shows cumulative density functions (cdfs) of *daysabs* separately for the treatment and control group for both pre- and post-reform periods using the full sample. All parts of the cdf of the treated clearly shifted to the left in the post-reform period as compared to the pre-reform period. In contrast to that, the cdf of the control group is very similar in both periods. Note that there is a clear shifting differential for up to 15 total

[15] Again I use a matched sample. I applied the same matching procedure as in Table 1.1 to the raw sample that—in this case—also includes respondents who had long-term absence spells. Out of the 21,211 observations, only 814 (3.8 percent) respresent a long-term spell. However, it is important to drop respondents in *all* of the years observed, even if they solely had a long-term spell in one of them. Without doing this, one might get biased estimates since it might be that the cut in long-term sick pay affected long-term absenteeism. However, this is the reason why the full sample has 21,211 obervations while my working sample in Approach I has only 17,803 although only 814 long-term spells were reported.

Table 1.4: Robustness Checks and Effect Heterogeneity: Intention-to-Treat Approach Using Matched Sample

Panel A: Robustness checks

	no job changers	**weighted + no refresh. sample**		**Yearly reform effects**	**effect on long-term absences**
DiD	0.0288*	0.0557***	DiD98	0.0329*	-0.0026
	(0.0161)	(0.0185)		(0.0206)	(0.0055)
			DiD97	0.0206	
				(0.0176)	
			DiD96	0.0312*	
				(0.0181)	
N	15,538	15,422		20,994	21,211

Panel B: Effect heterogeneity

	work conflict	**easy to find job**	**no probation**	**no fringe benefits**	**small firm**
DiD	0.0940**	0.0665*	0.0311**	0.0510**	0.0890***
	(0.0408)	(0.0351)	(0.0155)	(0.0246)	(0.0287)
Mean treated	0.4818	0.5160	0.5275	0.6044	0.5727
Change in %	+19.5	+12,9	+5.9	+8.4	+15.5
N	2,544	3,554	16,769	5,813	4,669

* $p<0.1$, ** $p<0.05$, *** $p<0.01$; standard errors in parentheses are adjusted for clustering on personal identifiers. All models make use of a matched sample and compare those who were intended to be treated—private sector employees—with those who were entirely unaffected by the reform (Approach I; group A vs. B in Figure 1.2). All models estimate OLS-DiD models as in equation (1.1). Every column in every panel represents one model. The dependent variable in all models except for the last column of Panel A is *noabs*. Column (4) in panel A estimates yearly treatment effects. In contrast to the standard models, it includes the semi-treated year 1996. *DiD96 (DiD97, DiD98)* is an interaction term between *T1* and *Year 1996 (Year 1997, Year 1998)*. Column (5) of Panel A uses the incidence of long-term absenteeism, i.e., a continuous spell of more than six weeks, as dependent variable. Long-term sick pay was cut from 80 to 70 % of foregone gross wages from January 1, 1997 onwards. In contrast to all other models in the chapter, column (5) of Panel A uses a sample that includes respondents with long-term sick leave in at least one of the years under consideration. *Work conflict* was only sampled in 1995. Thus, job changers are excluded in this model and only respondents with the same job as in 1995 are kept. *Work conflict* has a one for respondents who claimed to have conflicts with their boss. *Easy to find job* has a one for those who claimed that it would be easy to find an equivalent job in case of getting laid off. *No probation* has a one for those who are out of their probation period. *No fringe benefits* has a one for employees who received less than an annual total of €500 as fringe benefits.

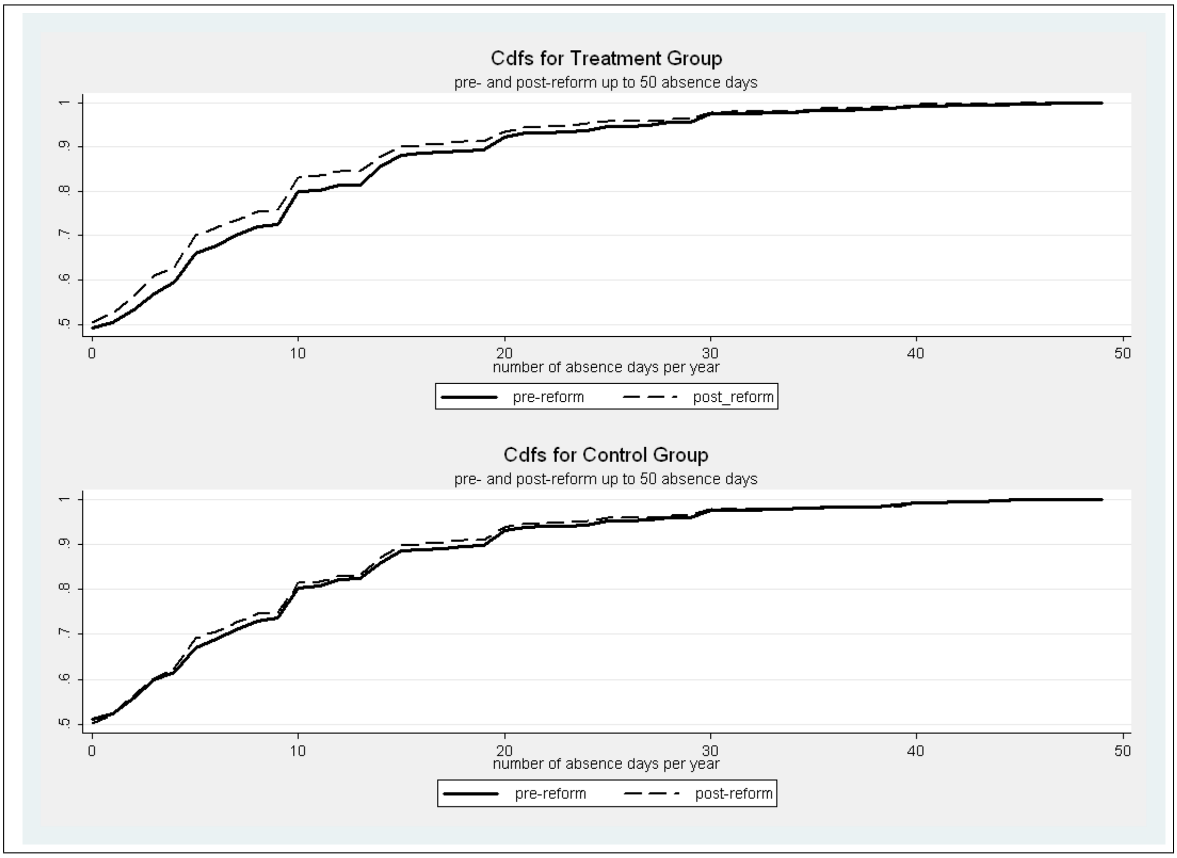
absence days. From 30 total absence days onwards, we do not observe any substantial shifts—neither for the treatment nor the control group. Please note that the cdf-shifts in Figure 1.5 are in line with my finding that employees primarily changed their short-term sick leave behavior.

Panel B of Table 1.4 tests effect heterogeneity. All models use *noabs* as dependent variable; thus the reference model is the one in column (4) of Table 1.2, where we found that the reform increased the share of employees with zero absence days by between six and eight percent. Column (1) uses only employees who state that they often have conflicts with their boss. Column (2) selects on employees who claimed that it would not be difficult to find an equivalent job if they were laid off. Both estimates are significantly different from zero and substantially larger than the baseline estimates, both absolutely and relatively, suggesting that these subsamples reacted stronger to the cut in sick pay. An explanation would be that employees with work conflicts have a poor intrinsic motivation and react stronger to monetary incentives. Employees who find jobs easily have better outside options which decreases the costs of getting laid off in case of shirking. In contrast to that, employees who are still in their probation period face a higher risk of getting laid off in case of shirking behavior. Consequently, as column (3) shows, they revealed weaker behavioral reactions than the subsamples in columns (1) and (2).
The last two columns of Panel B yield additional evidence on the credibility of my identification strategy and reinforces it. In column (4) I focus on employees who receive less than a total of € 500 per year as fringe benefits.[16] These employees were very likely to be affected by the cut in statutory sick pay. The result from this model supports this presumption since it yields a highly significant increase in the share on non-absent employees by 8.4 percent. A similar argument holds for the last column of Table 1.4: given the political economy of the reform (see Section 1.2.2), *a priori*, one would assume that employees in small firms showed a stronger behavioral reaction since employees in small firms are much less likely to be covered by collective agreements which might have codified voluntary sick pay on top of statutory sick pay. This is exactly what I find in column (5). The effect is significant at the 0.1 percent level and double the size in relative terms as compared to the baseline result.

As discussed above, an indirect method to test the common time trend assump-

[16] Total fringe benefits is the sum of all annual fringe benefits such as 13 and 14 month salaries, vacation bonuses, profit sharing and other bonuses like additional sick pay.

Figure 1.5: Cdfs for Treatment and Control Group Using Full Sample: Pre- vs. Post-Reform Periods



tion 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 1.5 demonstrates. Not only are the four estimated effects for the pseudo-reform years 1994 and 1995 insignificant and very close to zero: they also have reversed signs as compared to the models for the true reform year and the effects are precisely estimated. A simple power calculation with reference to the previously estimated treatment effects (i.e. 0.0337 for *noabs*) reveals that for a significance level of 0.05, the statistical power would be 0.86 (0.75) for the pseudo-reform year 1994 (1995). As above, the *daysabs*-model is imprecisely estimated. In conclusion, the main identifying assumption appears to be valid.

Table 1.5: Placebo estimates Using Matched Sample

	noabs	***daysabs***
T1×post1994	-0.0117	0.1467
	(0.0176)	(0.3639)
T1×post1995	-0.0025	0.0402
	(0.0159)	(0.3437)
Job characteristics	yes	yes
Educational characteristics	yes	yes
Personal characteristics	yes	yes
Regional unemployment rate	yes	yes
Time dummies	yes	yes
State dummies	yes	yes
N	13,467	13,467

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on personal identifiers. Each cell represents one OLS-DiD model. All models assign employees to treatment and control groups according to the same criteria as in the tables above. For example, I use a matched sample generated by the same procedure as the matched sample in Table 1.1. All models make use of a sample with 13,467 observations and information relying on the years 1993 to 1996. The interaction term between *T1* and *post1994 (post1995)* estimates pseudo-reform effects for the pseudo-reform year 1994 (1995). The dependent variable in column (1) is *noabs* and in column (2) *daysabs* (see section 1.3.1).

1.5.2 Specific Approaches II and III

Part B of the empirical analysis intends to disentangle specific reform effects from the overall effects in Part A. The two approaches focus solely on employees in the private sector and make use of alternative treatment and control groups. Since I just compare private sector employees—and not different occupational groups like in

Table 1.6: Specific Approaches II and III Using only Private Sector Employees (No Job Changers)

	Approach II: ***Cut first 3 days (A.2.1) vs. vs. unaffected (A.2.2)***		**Approach III:** ***No collective agreement (A.1) vs. vs. unaffected (A.2.2)***		**Robustness check:** ***Unaffected (A.2.2) vs. vs. unaffected (B)***	
	noabs	***daysabs***	***noabs***	***daysabs***	***noabs***	***daysabs***
DiD	0.1122***	-1.6482***	0.0796***	-1.5214***	-0.0416	0.7106
	(0.0331)	(0.4983)	(0.0296)	(0.5063)	(0.0412)	(0.9369)
Job characteristics	yes	yes	yes	yes	yes	yes
Educational characteristics	yes	yes	yes	yes	yes	yes
Personal characteristics	yes	yes	yes	yes	yes	yes
Regional unemployment rate	yes	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes
N	1,466	1,466	3,428	3,428	1,493	1,493
R^2	0.0755	0.0981	0.0673	0.0873	0.0813	0.0946

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are clustered on the industry×year level. Every column stands for one OLS-DiD model. Approaches II and III only use private sector employees and hence no matching prior to the regression analysis is performed. Approach II uses employees who are covered by collective wage agreements that implemented a cut in sick pay from 100 to 80 percent for the first three days of a sickness spell (A.2.1 in Figure 1.2). They are contrasted with private sector employees who are covered by collective wage agreements that kept the sick leave scheme unchanged (A.2.2 in Figure 1.2). Approach III contrasts private sector employees who are not covered by collective wage agreements (A.1 in Figure 1.2) with private sector employees who are covered by agreements that kept the replacement level unchanged (A.2.2 in Figure 1.2). Information on whether employees are covered by collective wage agreements was only sampled in the year prior to the reform, in 1995. Thus, only employees who have the same job as in 1995 are used. The last two columns display a falsification test which compares the two untreated groups A.2.2 and B (see Figure 1.2). Since this test again compares different occupational groups, i.e., private with public sector employees, prior to the regression analysis, every respondent in group A.2.2 is assigned a similar respondent out of the much larger pool of respondents in group B by means of one-to-one nearest neighbor matching. The dependent variable in columns (1), (3), and (5) is *noabs* and in columns (2), (4), and (6) it is *daysabs* (see section 1.3.1).

Approach I—it is not necessity to apply a matching procedure to generate the working sample. Approach II measures the effect for employees who unambiguously experienced a cut in sick pay from 100 to 80 percent of foregone gross wages for the first three days of a sickness episode. It compares group A.2.1 to group A.2.2 (see Figure 1.2 and Section 1.2.2, 1.3.2). Approach III contrasts employees who were not covered by collective agreements (A.1)—and who were thus likely to be effectively affected by the cut in statutory sick pay—with unaffected private sector employees in group A.2.2. All models rely solely on employees who held the same job as in the pre-reform year 1995 (see Section 1.3.2) and are based on OLS estimation.

Table 1.6 shows the results for both models and my two outcome measures. All four estimates show results that are significant at the one percent level. Moreover, all results have the correct sign, i.e., we find positive effects on the share of non-absent employees and negative effects on the average number of absence days. As expected, the effects are much stronger than the estimates for the intention-to-treat model. What is striking is that both models yield very similar results although they are based upon different treatment groups. Besides adding to the credibility of the identifed effects, this also indirectly supports my findings from above which suggest that short-term spells were primarily reduced as a result of the cut in sick pay (remember that the treatment group in Approach II only experienced sick pay cuts for the first three days of an episode). According to Approach II, the share of non-absent employees increased by 22 percent whereas the increase is about 15 percent in Approach III. Concerning the effect on the number of absence days, I find absolute decreases of 1.5 days which translate into relative decreases of 28 percent for Approach II and 31 percent for Approach III. Related to the 20 percent cut in the replacement level, these findings imply an arc elasticity—based on the average of replacement rates before and after the reform—of absence days that is larger than 1. With respect to the share of non-absent employees, the implied arc elasticity with respect to the replacement rate lies between 0.6 and 0.8.

In the last two columns of Table 1.6, I present an additional falsification check in which I compare the two untreated groups A.2.2 and B. As can be seen, the estimated effects are not statistically significant from zero.

Interestingly, the estimated reform effects from approaches I, II, and III also fit to the general reform setting, which is related to the organization of the German labor market and the political economy of the reform. We know that between a half

Table 1.7: Robustness Checks on the Common Group Error Structure: Donald-Lang

	Approach I		***Approach II***		***Approach III***	
	noabs	***daysabs***	***noabs***	***daysabs***	***noabs***	***daysabs***
DiD	0.0337**	-0.2037	0.1149**	-1.6193*	0.0787**	-1.6433***
	(0.0148)	(0.2826)	(0.0590)	(1.0965)	(0.0415)	(0.6619)
	[0.054]	[0.262]	[0.031]	[0.075]	[0.029]	[0.007]
Job covariates	yes	yes	yes	yes	yes	yes
Educational covariates	yes	yes	yes	yes	yes	yes
Personal covariates	yes	yes	yes	yes	yes	yes
Regional unempl. rate	yes	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes
N	8	8	34	34	197	197

* p<0.1, ** p<0.05, *** p<0.01; standard errors are in parentheses; p-values are in square brackets. Every column stands for one OLS-DiD model. The models are the same as in the tables above. For example, the model in column (1) is the same as the model in column (4) of Table 1.2 and the one in column (3) equals the one in column (1) of Table 1.6. To allow for the presence of unobserved common group errors, estimation was carried out by means of the Donald and Lang (2007) two-step procedure. Each regression is based on first-differenced aggregated residuals from the first stage. For Approach I, there are two groups (groups A and B in Figure 1.2) observed over five years, i.e., a total of 10 observations. For Approach II and Approach III, observations are aggregated at the industry level. There are in total 9 groups represented in Approach II and 65 in Approach III, but not all industries are represented in all years.

and two-thirds of all German employees were covered by collective agreements of which many guaranteed additional sick pay in form of fringe benefits. Although there are no official figures on this, a poll among craftsmens' businesses in the aftermaths of the reform suggests that around 50 percent of these companies did not apply the law (Brors and Thelen, 1998). Union representatives also mentioned this figure referring to the economy as a whole (Jahn, 1998). Thus it is appropriate to assume that about 50 percent of all employees in the private sector were directly affected by the reform. This figure fits quite well to the identified reform effects in my three models and to the results of the two specifications in columns (4) and (5) of panel B in Table 1.4. Considering that approaches I to III make use of three different treatment groups and two different control groups makes me confident that my models capture the reform effects in an appropriate manner.

In addition, one of the few official figures that is available is in line with my estimates. According to the German Ministry of Health (2009), the absence rate decreased in 1997-1998 as compared to 1993-1995 by about 15 percent.[17]

1.5.3 Robustness Checks on Common Group Errors and Placebo Estimates

As discussed in Section 1.4.2, I check whether the presence of unobserved common group errors might seriously affect my findings. The standard errors in Approach I were clustered on the individual level, while in Approach II and Approach III, the errors were routinely clustered on the industry$\times$year level. The actual reform implementation was, to a large extent, determined at this level.

Table 1.7 shows the results for all three approaches and the two outcome measures when I correct the standard errors for the presence of unobserved common group errors. I apply the Donald-Lang two-step method, which is a conservative correction method. Although the standard errors increase, in none of the six estimates do the coefficient becomes insignificant as a result of the correction method.

1.5.4 Reduction of Labor Costs and Job Creation

Since one of the main motivations of the reform was to reduce labor costs and to foster employment creation, Part C of the empirical analysis intends to assess the overall reform effect on labor costs. With this figure at hand, we can borrow findings from other studies that used general macroeconomic equilibrium models for Germany in the mid-1990s. This allows us to roughly calculate the (potential) job creation effect and give a better understanding to the dimensions of the reform.

First, I assess how the decreased obligation to provide sick leave benefits might have affected labor costs directly and indirectly. For the moment, we assume the world to be static. Then, the maximum overall decrease in labor costs can easily be calculated by comparing the total employer-provided sick pay benefit sum in the pre-reform years with the total benefit sum in the post-reform years with the assumption

[17] The absence rate is defined as the number of employees, on the first of a month, with a certified sickness divided by all employees. The cited figure only includes employees who were compulsorily insured with the Statutory Health Insurance, which includes the majority of Germans. However, the figure underestimates the true effect since public sector employees are included and sick leave without certificate (i.e. the first three days) is not counted.

that every employer strictly implemented the new law.[18] Thus, I calculate annual sick leave benefits for every employee in the sample and frequency weight this sum. For the pre-reform periods, I assume a replacement level of 100 percent of foregone gross wages and for the post-reform periods, I assume a replacement level of 80 percent. The frequency weighted benefit sums for both periods are multiplied with the frequency-weighted number of treated employees in my sample. By taking the difference between pre- and post-reform years, we obtain a total maximum decrease in labor costs of €2.8 billion per post-reform year.

This total amount of labor cost savings can be decomposed into three components. The first component is rooted in the lowering of the level of statutory sick pay from 100 to 80 percent of foregone gross wages. This amount is approximated by comparing the total sick leave payments in the pre-reform period to hypothetical sick leave payments for the same period and individuals assuming that the sick pay was already lowered at that time. I thus disentangle the direct savings effect from the savings effect that is induced by decreasing absence rates as a consequence of the reform. My estimates yield a maximum direct saving effect of €2.1 billion per year. If I assume that only half of all companies applied the new law stringently, these direct savings reduce to €1.05 billion.[19]

The second component represents the indirect labor cost effect that was triggered by the reform-induced decrease in workplace absences. From Table 1.3, we infer that the reform-induced reduction in absence days equaled about 0.6 days for employees with up to 5.5 annual absence days. Hence, using this subsample of employees, I take 0.6 times the average daily gross wage in the pre-reform years and multiply it by the frequency-weighted number of employees in these years. Thereby we obtain an indirect labor cost decrease of €315 million per year.[20] The residual is the third component, which is caused by time trends, changes in wages, and changes in the employment structure. The total reform-induced decrease in labor costs is thus $1.05 + 0.315 =$ €1,365 billion per year.

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

[19] I thereby implicitly assume that employees who worked in companies which applied the new law stringently did not differ systematically in terms of absence days and wages from those who worked in companies which voluntarily provided the old sick pay.

[20] Here, I use the same data set that I used to obtain the reform effect.

I crosscheck the plausibility of my labor cost calculations by looking at administrative data. The German Federal Statistical Office provides administrative data on the total sum of employer-provided sick pay for the whole of Germany, including voluntary sick pay. According to the German Federal Statistical Office (1995, 1997, 2000), the average total sick pay sum from 1993 to 1995 was €25.2 billion per year and decreased by €1.7 billion or 6.7 percent to €23.5 billion per year in 1997 and 1998.[21] This demonstrates the effectiveness of the reform. Note that my estimate of €1.4 billion is net of time trends and assumes that 50 percent of all private-sector employees were actually treated. On the one hand, the striking similarity of my 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 my identification strategy and the assumption that about 50 percent of all private-sector employees were affected by the reform.

In the very last step, I borrow findings from other macroeconomic studies—which were conducted at the time of the reform for Germany—to obtain a rough estimate of the potential job creation effect induced by decreasing labor costs. At that time, the majority of economists claimed that high labor were the main barrier for job creation in Germany. Indeed, all three available simulation studies, which are based on general equilibrium models, concordantly found a strong link between labor costs and employment creation (Zika, 1997; Meinhardt and Zwiener, 2005; Feil et al., 2008). Depending on the underlying assumptions, it seems reasonable to conclude that between 30,000 and 70,000 extra jobs could have been created in the long run because of the lower labor costs resulting from the reform —on the assumption of moderate short-term strike costs and a constant labor productivity.[22] Had the reform been accepted by employees and unions as fair-minded and had it been implemented strictly by all employers, twice as many jobs could have been created.

However, these simple calculations assume there were no general equilibrium responses to the reform. Furthermore, as the reforms led to mass demonstrations and strikes, the reduction in sick leave payments should be contrasted with the costs arising from this by-product of the reform. The notion that the reform did not predominately reduce moral hazard but induced more presenteeism and led to an overall

[21] 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.

[22] This figure needs to be interpreted relative to about 27.5 million dependent employees in the private sector at that time (German Federal Statistical Office, 1998).

drop in labor productivity should also be taken into consideration.

1.6 Conclusion

A natural experiment enables me to estimate the causal reform effect of a cut in the statutory sick pay level on sickness absence behavior and labor costs in Germany. I do this by relying on a conventional difference-in-differences methodology and three different estimation approaches. Typical selection issues common to evaluation studies are dealt with by employing longitudinal SOEP household data and thus identifying job changers who are the only ones who could have selected themselves out of the treatment. Moreover, I prove the robustness of my results in a number of checks and show why the common time trend assumption is likely to hold in this setting.

The cut in statutory sick pay from 100 to 80 percent of forgone gross wages intended to apply universally to every dependent employee in the private sector and was passed at the federal level. My first estimation approach intends to measure the overall reform effect among all employees in the private sector—the reform's target group. However, the non-acceptance of the reform by the population, which was manifested in mass demonstrations and union pressure, forced employers in select industries to voluntarily agree to the continuation of the old sick pay regime. Thus, my estimation approaches two and three exploit variation across industries and collective bargaining coverage.

My empirical findings suggest that the reform increased the overall share of private sector employees without any absence days by between six and eight percent. Looking at the impact on the average number of short-term absence days, quantile regressions reveal that the reform reduced this figure by around 12 percent for employees with up to 5.5 annual absence days. I clearly find that primarily employees in the lower tail of the sick leave days distribution adapted their sick leave behavior. This illustrates that moral hazard is of substantial relevance in this part of the distribution.

Estimates suggest that about half of all employees in the private sector effectively experienced cuts in their sick pay. Exploiting differences in the reform implementation at the industry level, my estimation approaches two and three indeed show that a strict enforcement of the cut in statutory sick pay increased the share of non-absent employees by between 15 and 20 percent and decreased the number of absence days by about 30 percent. The implied arc elasticity of total absence days with respect to the replacement level in thus larger than one, whereas the arc elasticity of being non-absent with respect to the replacement level is about 0.8. The true elasticities are severely underestimated when solely relying on the overall reform effects. The calculated elasticities are comparable to the elasticities found in Johansson and Palme (2005).

I estimate that the direct labor cost savings effect due to the decrease in benefit levels clearly exceeded the indirect savings effect due to the decrease in absenteeism. Official data on employer-provided sick pay show that the total sick pay sum decreased by 6.7 percent or €1.7 billion per year in the post-reform as compared to the pre-reform years. Given that this figure includes all kinds of time trends, it is very consistent with my estimate which is based on SOEP data and yields an annual savings effect of €1.4 billion.

Using the findings of various other studies which are derived from macroeconomic simulation models for Germany, back-of-the-envelope calculations suggest that the reform might have led to the creation of between 30,000 and 70,000 new jobs. Had the reform been implemented strictly by all companies, as was intended by the policymakers, the job creation effect could have been double this size. However, possible general equilibrium effects of the reform and unintended side-effects such as strikes and mass demonstrations are beyond the scope of this chapter and may have offset or even overcompensated the pure reform effects.

Germany is the country of origin of Bismarckian corporatism adopted by almost all European countries. The organization of the labor market heavily relies on social partnership and autonomy in bargaining with a bargaining coverage exceeding 50 percent of all employees. Thus, the evaluation of this reform also illustrates how reform intention and actual reform implementation may diverge when labor market reforms are planned and passed on the federal level while collective bargaining dominates the organization of the labor market on the lower level. This, in turn, leads me to the conclusion that policymakers should at least improve their way of communicating such reforms. To what extent the success of such reforms depends on cultural peculiarities and macroeconomic conditions is of importance and further studies on this subject would be valuable.

Appendix A

Table 1.8: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.	Obs.
Dependent variables					
Noabs	0.525	0.499	0	1	20,700
Daysabs	5.172	9.274	0	365	20,700
Personal characteristics					
Female	0.437	0.496	0	1	20,700
Age	39.05	11.01	18	65	20,700
Age squared	1,646	888	324	4,225	20,700
Immigrant (=1 if immigrant)	0.16	0.367	0	1	20,700
East (=1 if residing in East Germany)	0.263	0.44	0	1	20,700
Partner (=1 if partner)	0.761	0.427	0	1	20,700
Married (=1 if married)	0.669	0.471	0	1	20,700
Children (=1 if children)	0.474	0.499	0	1	20,700
Disabled (=1 if disabled)	0.031	0.172	0	1	20,700
Health good (best 2 of 5 SAH categories)	0.669	0.471	0	1	20,700
Health bad (worst 2 of 5 SAH categories)	0.067	0.25	0	1	20,700
No sports (=1 if no exercise)	0.38	0.485	0	1	20,700
Educational characteristics					
Drop out	0.038	0.192	0	1	20,700
8 years of educational attainment	0.292	0.455	0	1	20,700
10 years of educational attainment	0.355	0.479	0	1	20,700
12 years of educational attainment	0.043	0.204	0	1	20,700
13 years of educational attainment	0.189	0.392	0	1	20,700
Other certificate	0.082	0.274	0	1	20,700
Work in job trained for	0.567	0.495	0	1	20,700
Job characteristics					
Part-time employed	0.139	0.346	0	1	20,700
New job	0.128	0.334	0	1	20,700
No. years in company	9.068	9.061	0	48.7	20,700
Small company	0.264	0.441	0	1	20,700
Medium company	0.267	0.442	0	1	20,700
Large company	0.218	0.413	0	1	20,700

Continued on next page...

... Table 1.8 continued

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**Obs.**
Very large company	0.22	0.414	0	1	20,700
Blue collar worker	0.296	0.457	0	1	20,700
White collar worker	0.489	0.5	0	1	20,700
Civil servant	0.07	0.256	0	1	20,700
Public sector	0.276	0.447	0	1	20,700
Self-employed	0.091	0.287	0	1	20,700
High job autonomy	0.257	0.437	0	1	20,700
Gross wage per month	2,030	1,401	0	51,129	20,700
State unemployment rate	11.565	3.962	6.3	21.7	20,700

Chapter 2

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

Abstract

This chapter analyzes the effects of an increase in statutory sick pay from 80 to 100 percent of forgone gross wages in Germany. Difference-in-differences approaches show that the increase in generosity decreased employee attendance by about ten percent or one day per employee per year. Heterogeneity in response behavior was of great importance and employee health its main driver. For employers, the increased contribution represented increased labor costs of about €1.8 billion per year. My empirical evidence supports the notion that employers tried to compensate for this shock to labor costs by increasing overtime and decreasing wages.

2.1 Introduction

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

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

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

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

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

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

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

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

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

2.2 The German Sickness Insurance System and Policy Reform

2.2.1 The Sick Pay Scheme and Monitoring System

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

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

Monitoring is carried out primarily by the *Medical Service of the SHI*. One of the main objectives of the Medical Service is to monitor sickness absence. German social legislation requires the SHI to contact the Medical Service and request a medical opinion to resolve any doubts regarding the validity of sick leave claims. Such doubts may arise if someone is absent for short periods with unusual frequency or is regularly sick on Mondays or Fridays. Similarly, if a doctor certifies sicknesses

[1] The entitlement is codified in the *Gesetz über 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.

with unusual frequency, the SHI may call for an expert assessment of that doctor. The employer also has the right to request an expert assessment by the Medical Service, which is based on medical records, workplace information, and a statement that the patient is required to submit. If necessary, the Medical Service has the right to conduct a physical examination of the patients and to cut their benefits.[2] In 2007, about 2,000 full-time equivalent and independent doctors worked for the medical service and examined 1.7 million cases of absenteeism (Medizinischer Dienst der Krankenversicherung (MDK), 2008).

2.2.2 The Policy Reform

In the election campaign of 1998, the Social Democrats and the Greens promised to increase statutory sick pay from 80 to 100 percent of foregone gross wages should they form a new coalition government. The announcement was a reaction to a cut in sick pay under the previous conservative government under Chancellor Kohl in October 1996. At that time, together with a reduction of long-term sick pay, short-term sick pay was decreased from 100 to 80 percent of forgone gross wages. Ziebarth and Karlsson (2009) analyze the effect of the cut in short-term sick pay and find that it reduced the average number of absence days by about 5 percent.[3] The majority of Germans perceived the cut in sick pay as unfair and socially unjust and a number of strikes were organized opposing it. Immediately after the election of the new center-left government in September 1998, a law was passed that went into effect on January 1, 1999, increasing statutory short-term sick pay from 80 to 100 percent of foregone gross wages.[4]

Although statutory sick pay was increased by 20 percentage points in 1999, it did not change conditions for all private-sector employees, since employers can voluntarily

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

[3] For various reasons, it makes sense to analyze the effects of a decrease in coverage separately from the effects of an increase in coverage: first, the effects may be expected to differ. Second and more importantly, the sick pay reform was accompanied by various other reforms that act as confounding factors in the estimation: for example, a waiting period for new hires was introduced, the basis of calculation was changed, and long-term sick pay was also cut. Moreover, treatment and control groups differ among all these reforms, which requires different identification strategies. In addition, Ziebarth and Karlsson (2009) only employ conventional parametric difference-in-differences models and do not provide evidence on either the underlying operating mechanisms or employers' responses to the reform. The estimated change in sick leave behavior is, however, in line with the findings of the present chapter.

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

provide sick pay over and above the minimum requirements. This, incidentally, is always a problem when analyzing the effects of changes in statutory minimum requirements, i.e, *all* studies in this strand of the literature face this issue. After the cut in sick pay in October 1996, and partly in response to union pressure, employers from various sectors had agreed to continue paying 100 percent of wages during sick leave in collective wage agreements. There are no official figures on how many employees benefited from this, but in 1998, union leaders proudly declared that 13 out of 27 million employees would receive 100 percent sick pay (Jahn, 1998).[5] In 1997, a poll among craftsmen's businesses showed that 51 percent were voluntarily providing 100 percent sick pay, probably due to the close relationship and mutual trust between employers and employees in these small companies (Ridinger, 1997). Like all of the other studies that have evaluated the impact of changes in statutory benefit levels, I 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.

Even with 80 percent statutory sick pay, Germany provides 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 was general consensus among German economists that the high overall labor costs were 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).

2.3 Data and Variable Definitions

For the empirical analysis, I 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

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

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, I exploit information on sick leave for the years 1997 to 2000.[6] I restrict my working sample to respondents who are employed full-time and between 25 and 55 years of age. I do not use respondents with item non-response on relevant variables.

2.3.1 Sick Leave Measure and Covariates

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

I 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 issue 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. Measurement errors inflate standard errors and lead to less precise estimates. They would seriously hamper my analysis if the reform had any impact on them. This is very unlikely since the reform of statutory sickness insurance system was the subject of heated political and media debate during the entire period under consideration. Moreover, in the case of underreporting, with $0 < \alpha < 1$ being the reported fraction of true days, the estimate would be downward biased by α.

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

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

doctor-certified sick leave is included, and that employer-provided sick leave is often left out. This almost always leaves the researcher with censored data and makes certain types of analyses impossible.[7] However, having an uncensored measure of the total number of days of sick leave based on survey data comes at the cost of not having detailed spell data.

The whole set of explanatory variables can be found in Table 2.1. The control variables used the main specifications (Part A of Table 2.1) 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*, I 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.

For the extended analyses in the second part of the chapter, I employ additional covariates (Part B of Table 2.1). These additional covariates either incorporate a substantial degree of item-non-response or were only collected in specific years, which is why I do not use them in the main specifications. Further details about these variables can also be found in the notes to Table 2.6.

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

Table 2.1: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Treatment Group	0.6566	0.4749	0	1	23,058
Daysabs	9.7793	24.9027	0	365	23,058
A: Variables used in main specifications					
Personal characteristics					
Female	0.3188	0.466	0	1	23,058
Age	39.69	8.22	25	55	23,058
Age squared	1643	661	625	3025	23,058
Immigrant	0.1433	0.3504	0	1	23,058
East German	0.2654	0.4415	0	1	23,058
Partner	0.7944	0.4042	0	1	23,058
Married	0.6803	0.4664	0	1	23,058
Children	0.4795	0.4996	0	1	23,058
Disabled	0.0438	0.2048	0	1	23,058
Health good (best two of five SAH categ.)	0.6355	0.4813	0	1	23,058
Health bad (worst two of five SAH categ.)	0.0811	0.2731	0	1	23,058
No sports	0.3776	0.4848	0	1	23,058
Educational characteristics					
Drop out	0.0269	0.1618	0	1	23058
Degree after 8 years' schooling	0.2883	0.453	0	1	23,058
Degree after 10 years' schooling	0.3659	0.4817	0	1	23,058
Degree after 12 years' schooling	0.0512	0.2204	0	1	23,058
Degree after 13 years' schooling	0.1929	0.3946	0	1	23,058
Other degree	0.0746	0.2627	0	1	23,058
Years with company	9.5069	8.4721	0	41	23,058
Trained for job	0.6063	0.4886	0	1	23,058
Job characteristics					
New job	0.1573	0.3641	0	1	23,058
Blue-collar worker	0.3707	0.483	0	1	23,058
White-collar worker	0.4602	0.4984	0	1	23,058
One man company	0.0328	0.1781	0	1	23,058
Small size company	0.2441	0.4296	0	1	23,058
Medium size company	0.2754	0.4467	0	1	23,058
Large company	0.2211	0.415	0	1	23058
Very large company	0.2266	0.4186	0	1	23058
Gross wage per month	2467	1278	404	28632	23,058
Annual state unemployment rate	11.36	4.53	5.4	21.7	23,058

Continued on next page...

... *Table 2.1 continued*

Variable	Mean	Std. Dev.	Min.	Max.	N
B: Variables used in extended analyses					
Long-term absence	0.0558	0.2296	0	1	23,058
Job loss	0.1102	0.3131	0	1	23,058
Not impaired by health	0.7826	0.4125	0	1	23,058
Severely impaired by health	0.0236	0.1518	0	1	23,058
Low health satisfaction (0-4 on scale 0/10)	0.0528	0.2237	0	1	23,058
High health satisfaction (10 on scale 0/10)	0.0528	0.2237	0	1	23,058
Overtime hours per week	2.6811	3.8891	0	23.1	20,732
Gross wage last month (job change prev. yr.)	2170	1135	409	12,782	3627
Low job satisfaction (0-4 on scale 0/10)	0.055	0.2281	0	1	22,347
Very worried about job security	0.1566	0.3634	0	1	22,503
Job makes no fun ('97; no job changers)	0.1301	0.3365	0	1	12,786
Not religious ('97)	0.3649	0.4814	0	1	15,537
Sickness should be insured by state ('97)	0.3804	0.4855	0	1	15,537
Sickness should be insured privately ('97)	0.0857	0.2799	0	1	15,537
Expects job loss within 2 years ('98; no job changers)	0.0875	0.2826	0	1	16,771
Expects promotion within 2 years ('98; no job changers)	0.1958	0.3968	0	1	16,771
Firm reduced workforce last year ('99; no job changers)	0.2795	0.4488	0	1	17,201
Control life ('99)	0.2888	0.4532	0	1	17,351
Can influence life ('99)	0.4083	0.4915	0	1	17,351
Need to work hard for success ('99)	0.5257	0.4994	0	1	17,351
No work council in firm ('01)	0.3841	0.4864	0	1	16,161

Variables with years in parenthesis were only surveyed in the corresponding year. When the information sampled refers to the workplace, only respondents who still work at the same workplace are kept. For example, respondents who answered the work council question in 2001 are kept in all years in which they were interviewed and worked at the same workplace as in 2001. For variables that were only surveyed in one year but do not contain workplace information, I 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.

2.3.2 Treatment and Control Group

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

2.4 Estimation Strategy and Identification

2.4.1 Assessing the Causal Reform Effects on Sickness Absence

Parametric Approaches

OLS

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

$$y_{it} = \lambda p99_t + \pi D_{it} + \theta DiD_{it} + s'_{it}\psi + \rho_t + \phi_s + \epsilon_{it} \tag{2.1}$$

where y_{it} stands for the annual number of days of sick leave for individual i in year t, $p99_t$ is a post-reform dummy, D_{it} is the treatment group dummy, and DiD_{it} is the regressor of interest. It has a one for respondents in the treatment group in post-reform years and gives us the causal reform effect should certain assumptions hold true. It can also be interpreted as the interaction term between the treatment group dummy and the post-reform dummy. By including additional time dummies ρ_t I control for common time shocks that might affect sick leave. State dummies ϕ_s account for permanent differences across the 16 German states along with the annual state unemployment rate that controls for changes in the tightness of the regional labor market and that is included in the $K \times 1$ column vector s'_{it}. The other $K-1$ regressors are made up of personal controls including health status, educational controls, and job-related controls as explained in Section

2.3.1. In addition to the covariates that are displayed in Part A of Table 2.1, I 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 (2.1) 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, I found the so-called *Zero-Inflated Negative Binominal Model (NegBin)* to be appropriate for my purposes.

The underlying statistical process differentiates between absent employees and non-absent employees and assigns different probabilities, which are parameterized as functions of the covariates, to each group. The binary process is specified in form of a logit model and the count process is modeled as an untruncated NegBin-2 model for the binary process to take on value one. Hence, zero counts may be generated in two ways: as realizations of the binary process and as realizations of the count process when the binary process is one (Winkelmann, 2008). In contrast to the more restrictive Poisson distribution, the employed negative binomial distribution not only takes excess zeros into account but also allows for overdispersion and unobserved heterogeneity.[8] In the notation of Cameron and Trivedi (2005), the NegBin distribution can be described as a density mixture of the following form:

$$\begin{aligned}
\varphi(y|\mu,\alpha) &= \int f\left(y|\mu,\nu\right) \times \gamma(\nu|\alpha)\, d\nu \\
&= \int_0^\infty \left(\frac{e^{-\mu\nu}\{\mu\nu\}^y}{y!}\right)\left(\frac{\nu^{\delta-1}e^{-\nu\delta}\delta^\delta}{\Gamma(\delta)}\right) d\nu \\
&= \frac{\Gamma(\alpha^{-1}+y)}{\Gamma(\alpha^{-1})\Gamma(y+1)}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\mu+\alpha^{-1}}\right)^{y} \qquad (2.2)
\end{aligned}$$

[8] The unobserved heterogeneity allowed for in the NegBin-2 is based on functional form and does not capture unobserved heterogeneity, which is correlated with explanatory variables.

where $f(y|\mu,\nu)$ is the conditional Poisson distribution and $\gamma(\nu|\alpha)$ is assumed to be gamma distributed with ν as an unobserved parameter with variance $\alpha = 1/\delta$. Note that in the special case of $\alpha = 0$ the NegBin collapses to a simple Poisson model. $\Gamma(.)$ denotes the gamma integral and $\mu = exp(x_{it}'\beta)$ with x_{it}', incorporating all the regressors as in equation (2.1).

The marginal effect of the interaction term DiD_{it} is—given that the model assumptions are fulfilled—the causal reform effect and in the following is always displayed together with output tables.[9]

Non-Parametric Approaches

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

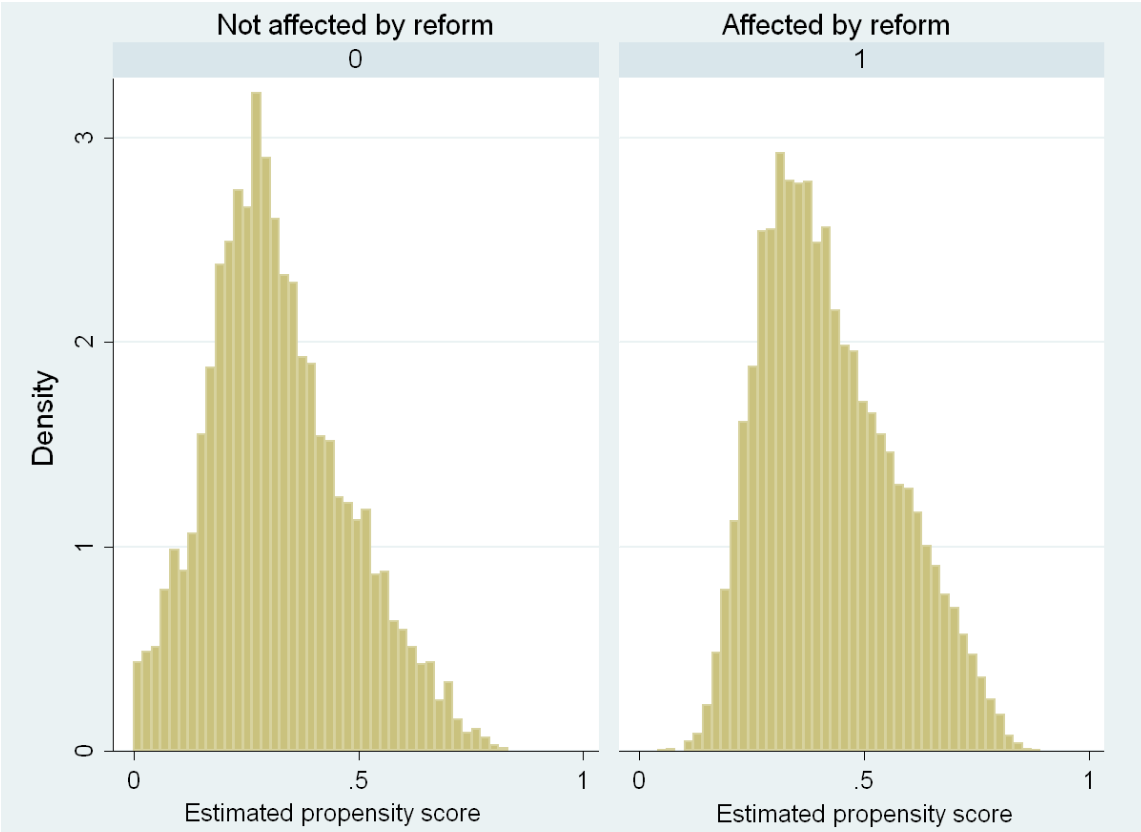
Most matching analyses use the method of propensity score (PS) matching. Rosenbaum and Rubin (1983) have shown that conditioning on the propensity score (PS)—the probability of being selected into the treatment group—is equivalent to selecting pairs of treated and control observations based on every covariate dimension, provided that unconfoundedness holds. I estimate the PS by means of a logit model and select the covariates to be included out of the total number of covariates (Part A of Table 2.1) using likelihood ratio tests on zero coefficients. In a first step, I do this for control variables in levels and in a second step for their interactions (Imbens, 2008).

In addition to a plausible selection on observables, matching requires that the distributions of the covariates for treated and control observations overlap to a large

[9] Puhani (2008) has shown that the advice of Ai and Norton (2004) to compute the discrete double difference is not of relevance in nonlinear difference-in-differences models when the interest lies in the estimation of a treatment effect. The average treatment effect on the treated at the time of the treatment is given by $\varphi(y|\alpha, \bar{s}_1, p99 = 1, D = 1, DiD = 1) - \varphi(y|\alpha, \bar{s}_1, p99 = 1, D = 1, DiD = 0)$, where $\bar{s}_1$ denotes the average values of the covariates for the treatment group in the post-treatment period. This is exactly what I calculate and present throughout this chapter.

extent. In this setting, the common support assumption is fulfilled, as can be seen in Figure 1. The PS distribution for both groups shows a large overlap with the region of common support lying between PS values of 0.05 and 0.92.

Figure 2.1: Distribution of Propensity Scores Showing Region of Common Support



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

The second method is k-to-one nearest neighbors matching with replacement.

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

Combining Parametric and Non-Parametric Approaches

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

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

$$\Delta s = \frac{\bar{s}_1 - \bar{s}_0}{\sqrt{\sigma_1^2 + \sigma_0^2}} \tag{2.3}$$

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

Applied to my case, I first look at how the covariate distribution for the treatment group differs in comparison to the control group, i.e., I compare private-sector employees to those whose sick pay was not affected throughout the whole period under consideration. Table 1.1 shows in column (1) the means of the covariates for the treatment group and in column (2) the means of the covariates for the control group. It appears that the two groups are very similar with respect to their observable characteristics. This presumption is reinforced by column (3), which displays the normalized difference. Indeed, all of the values are smaller than 0.20 and some tend towards zero.

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

Table 2.2: Sample Means of Treatment and Control Group: Raw, Matched, and Blocked Sample

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

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

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

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

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

Imbens and Wooldridge (2009) propose two approaches that both combine the strengths of parametric and non-parametric estimators; both estimators work well in practice, and both estimators lead to robust results. Approach number one is a combined blocking and regression approach. In the first step, stratification matching is applied to the raw sample to obtain a blocked sample with better balancing properties, as in columns (7) to (9). In the second step, parametric regressions—as detailed in Section 2.4.1—are run within each block. Then, the within-block treatment effects are weighted by the relative size of the blocks and aggregated into an overall average treatment effect on the treated. The crucial point is that the covariate distributions within each stratum are very similar and, thus, out-of-sample predictions are avoided.

The second approach also aims to smooth differences in covariates between treatment and control group and additionally corrects for the bias described in Abadie and Imbens (2007). It combines regression and k-nearest neighbors matching. In the first step, using only the untreated who were matched to the treated, I conduct a linear regression of the outcome on the covariates.[10] Then, in the second step, the counterfactual potential outcome for the case without treatment, y^0, is calculated as (Abadie et al., 2004):

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

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

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

Identification of Causal Effects

In the previous subsections, I discussed how I utilized the rich set of socioeconomic background information to make the treatment and the control group as comparable as possible. As can be seen in columns (6) and (9) of Table 1.1, both matching and blocking yield two samples that are almost identical in terms of observables.

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

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

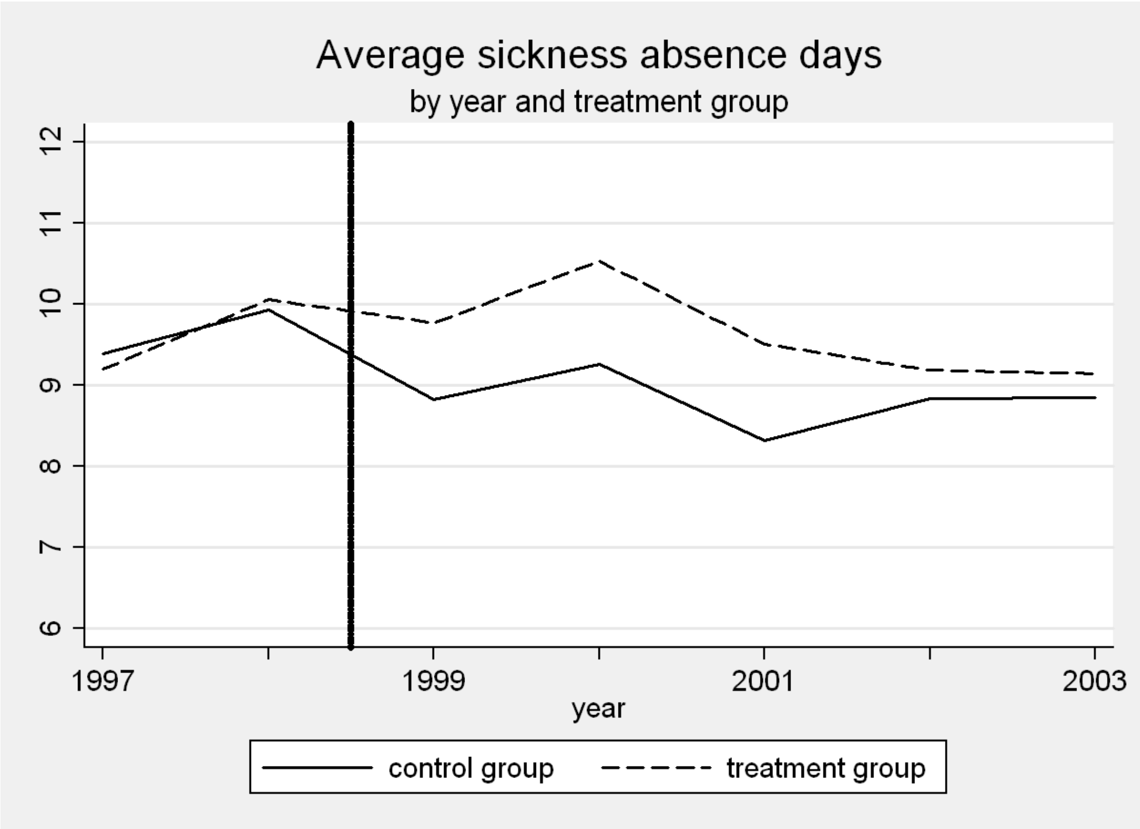
Second, I can exclude the possibility that selection into or out of the treatment contaminated my estimates since I rely on panel data and can identify job changers. For example, I can test the robustness of my results with respect to sample composition changes over time and labor market attrition. As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

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

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

Finally, in Figure 2, I display the average number of sickness absence days for several pre- and post-reform years and both groups. In 1996, as explained in Section 2.2, various sick leave reforms were implemented that all affected subsamples that differ from those analyzed here: thus, I can only use the pre-reform years of 1997 and 1998 for this exercise. However, I 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, and since the reform was hotly debated in the media, a priori, I 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 I find. I observe relatively parallel curves for both groups in the pre-reform years. After the reform went into effect in 1999, the absence curve for the treatment group shifts upwards and subsequently runs parallel to the curve for the control group. In this graph, it seems as if the reform effect lasts for about four years, since I then observe a closing of the gap in absence days. However, it should be kept in mind that Figure 2 paints a raw, unconditional picture. One explanation for the closing of the gap from 2001 to 2002 could be that unemployment rates increased by five percent or 0.5 percentage points (200,000 unemployed) between the two years—after they had been decreasing more or less monotonically for four years, i.e., since 1997 (German Federal Statistical Office, 2009a). It is a well-documented stylized fact that changes in unemployment and absence rates are negatively correlated (Askildsen et al., 2005). While the figure here represents only descriptive evidence, I also correct the sample composition with respect to a rich set of covariates in the empirical assessment below.

Figure 2.2: Average Sickness Absence Days for Treatment and Control Group over Time



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

As has been discussed in Section 2.2.2, even before the increase in statutory sick pay, some employers agreed in collective bargaining to provide 100 percent sick pay voluntarily. I cannot precisely identify employees who were subject to such collective wage agreements. Since employers are always free to provide fringe benefits on top of statutory regulations, it is intrinsically difficult to identify all contractually fixed payments. The approach equals an intention-to-treat approach, which is always required when focusing on changes in statutory minimum standards: in this strand of the literature, the *overall* effects of changes in statutory sick pay are evaluated. In contrast to other countries like Sweden, where differences in the labor agreements are more fragmented, polls for Germany at the time of the reform suggest that around half of all private-sector employees received statutory (80 percent) sick pay and the other half received 100 percent sick pay (Ridinger, 1997; Jahn, 1998). I believe that my analysis illustrates clearly how changes in federal minimum standards translate into real-world effects in a labor market that is characterized by Bismarckian corporatism and where unions and employers' representatives negotiate over the level of fringe benefits and wages on the industry or firm level. Almost all European countries are characterized by this form of corporatism.

2.4.2 Assessing Heterogeneity and Further Reform Effects

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

We start by examining whether heterogeneity in the response to the policy change plays a role. A priori, one would expect that the reform effect is not uniformly distributed across socio-economic and workplace characteristics. For example, by differentiating the reform effects by the health status of the respondents, I provide evidence on whether changes in employee sick leave behavior is attributable primarily to shirking or to presenteeism. Technically, I assess treatment effect heterogeneity by interacting possible covariates with the regressor *DiD* in equation (2.1). Then, I add this additional interaction term to the model.

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

2.5 Empirical Results

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

2.5.1 Assessing the Causal Reform Effects on Sickness Absence

Parametric Approaches

I start by estimating parametric OLS and ZINB-2 models on the raw sample with all covariates of Part A of Table 2.1 included. In the following, I always display marginal effects. The parametric DiD estimates are displayed in columns (1) and (2) of Table 2.3. The OLS model yields an estimate of 1.366 that is statistically significant at the five percent level. The ZINB-2 model gives an estimate of 1.018 with a standard error of 0.468. The unconditional double difference of the means of the two groups for the two time periods is 1.441 (std. err. 0.732; not shown) and very close to the OLS estimate in column (1), which reinforces the credibility of the common time trend assumption.

Non-Parametric Approaches

Columns (3) and (4) give the results when two different matching variants are applied using the matched subsample of Table 1.1. In the first step, the matched subsample of Table 1.1 is obtained by five-to-one nearest neighbors matching on the probability of belonging to the treatment group vs. the control group. As can be seen in column (6) of Table 1.1, the covariate distribution is almost perfectly balanced between the treatment and control group in this sample. In the second step, I perform two matching methods as described in the methods section. In this respect I could also call my non-parametric applications "two-step matching procedures." While I conduct matching on the probability of belonging to the treatment group to obtain the matched sample, afterwards, I perform matching on the difference-in-differences indicator. In other words, in the first step, I match on the treatment group dummy D, and in the second step, I match on the difference-in-differences indicator DiD.

Hence, both columns (3) and (4) make use of the matched sample as shown in Table 1.1. In column (3), the underlying matched sample is stratified into blocks based on the propensity score (PS) that DiD=1. Then, by taking the average values of treated and untreated within each block, the block-specific reform effects are calculated, which are finally aggregated to a weighted overall average. This method gives an estimate of 1.138 with a standard error of 0.406. Column (4) yields the estimate when five-to-one nearest neighbors matching is applied to the difference-in-differences indicator using the matched sample. The estimated reform effect is 1.120 and significant at the one percent level.

Table 2.3: Difference-in-Differences Estimation: Parametric, Non-Parametric, and Combined Methods

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

* $p<0.1$, ** $p<0.05$, *** $p<0.01$; standard errors are in parentheses. In the parametric specifications, they are adjusted for intrapersonal correlations. The estimate in column (2) is the marginal effect, calculated at the means of the covariates except for the post reform dummy (=1), the treatment group dummy (=1), the year 1999 dummy (=1), the year 2000 dummy (=1), and *DiD* (=1). ZINB-2 stands for *Zero-Inflated Negative Binominal Model 2*. In columns (3) and (4), the "matched sample" of Table 2.2 is the underlying sample. In column (3), the propensity score (PS) of belonging to the treatment group in post-reform years (*DiD*=1) is estimated, based on a logit model and the same covariates as in Table 2.2. Based on this PS, the sample is stratified into eleven blocks, each with an equal PS for treated and non-treated. 119 observations (12 treated) are outside the region of common support. Then, the block-specific treatment effects – the difference in average outcomes for treated (*DiD*=1) and non-treated (*DiD*=0)– are weighted by the number of treated to obtain the overall average treatment effect on the treated. In column (4), the average treatment effect on the treated is obtained by five-to-one nearest neighbors matching. In that specification, 150 observations lie outside the common support (31 treated). Standard errors in column (4) are obtained by bootstrapping with 100 replications. In column (5), the "blocked sample" of Table 2.2 is used. Then within each block, a ZINB-2-DiD regression is performed. Finally, the within block estimates are weighted by the number of treated observations to obtain the overall treatment effect on the treated. In column (6), five-to-one nearest neighbors matching and regression are combined. As explained in Abadie et al. (2004), the estimator is bias corrected and allows for heteroskedastic errors. In all columns, except for columns (3) and (4), the number of treated observations is 7,199. In column (3) [(4)], the number of treated observations is 7,187 [7,168] since 12 [31] observations lie outside the region of common support.

Combining Parametric and Non-Parametric Approaches

According to Imbens and Wooldridge (2009), the most suitable methods combine regression and matching and are thus more flexible and robust than other methods. Column (5) shows the result when the raw sample is first stratified on the probability of belonging to the treatment group (hence it makes use of the blocked sample in Table 1.1) and then regressions as in equation (2.1) are run block-by-block. The overall treatment effect, which is 0.8973 and significantly different from zero, is obtained as an average of the within-block estimates weighted by the block size. The method used in the last column also combines matching and regression and eliminates a bias that has been proven to exist for nearest neighbor matching. More details can be found in Section 2.4.1 and Abadie and Imbens (2007). The resulting estimate is similar to the one in column (5) and yields a reform effect of 0.9986 (std. err. 0.33307).

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

If we take the mean number of absence days in the pre-reform period for the treatment group, which was 9.7, and relate an estimate of 0.97 to it, we would conclude that the increase in statutory sick pay led to a 10 percent increase in the average number of absence days among the treatment group. Since the average private-sector worker in my sample had a pre-reform gross wage of €2,272 per month, the daily statutory sick pay before the reform was €61 and after the reform €76. Note that the reform made workplace absences absolutely costless in monetary terms, and that employees on sick leave gained €15 on average as compared to the pre-reform period. Under the assumption that only half of all private-sector employees effectively experienced an increase in sick pay, my estimates would suggest that the employees affected increased their days of sick leave by about two per year. The implied elasticity with respect to the increase in the replacement rate would be 0.9. This finding is comparable with the results of the few existing studies that analyze similar reforms (Johansson and Palme, 2005).

Robustness Checks

Apart from having analyzed the sensitivity of the results with respect to various parametric, non-parametric as well as combined methods, further robustness checks are shown in Table 1.4. Column (1) displays the result for a model that includes the lagged level of the total number of absence days as an additional covariate. This specification yields a positive and highly significant reform estimate of 1.568.

Column (2) checks whether panel or labor market attrition might drive my results. Only those who were observed working in the pre- *and* post-reform period at least once are included in the sample. By restricting the sample as such, we lose approximately 8,000 observations, and the precision of the estimate decreases; the estimate is only significant at the 11.8 percent level. However, we find a reform effect that is of similar size to the one in the main specification in column (1) of Table 2.3. When applying the Abadie and Imbens (2007) nearest-neighbors matching regression method to that specification, precision increases substantially, and the effect is significant at the one percent level (result not shown).

Columns (3) and (4) deal with concerns that selection into occupations might produce or bias the results. Column (3) excludes respondents who answered the following question with "yes:" *"Did you change your job or start a new one after December 31, 199X?"* I thereby capture job changers who might have selected themselves into (or out of) the treatment. The size of the estimate is almost identical to the main estimate in the first column of Table 2.3 and is marginally significant at the ten percent level. This also holds for column (4), where I provide an alternative robustness check on selection effects. In column (4), all private-sector employees who changed industry sector in the post-reform period are excluded from the sample. Given the reform design, it is likely that collective bargaining assured that sick leave regulations only varied across but not within industries. In another check, we look at whether the reform was followed by a change in the rate of job change. There is no evidence that this occurred. Between January 1997 and the date of the 1998 SOEP interview (most are conducted during the first three months of the year), 16.45 percent of all interviewees had changed jobs. This rate remained almost exactly the same for the period between January 1999 and the 2000 SOEP interview, namely 15.78 percent. In addition, looking at whether the distribution of job-changers across health states changed after the reform provides no such evidence either. From 1997 until the 1998 interview, 14.43 percent of all employees in poor or bad health changed

Table 2.4: Robustness Checks

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

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers, except for column (5) where they are clustered on state × year (64 cluster) and column (6) where they are clustered on the industry × year (242 cluster) level. All specifications are as in column (1) of Table 2.3 except for the following: The model in column (1) contains the lagged number of annual absence days as an additional covariate. The model in column (2) includes only those who were observed at least once (working) in the pre-reform years *and* the post-reform years. The model in column (3) excludes all those who have changed their jobs in the year prior to the interview. The model in column (4) excludes private sector employees who have changed their industry branch in the post-reform period. The model in in column (7) estimates the reform effect on the incidence of long-term absenteeism, i.e., a sickness period of more than six weeks.

jobs. From 1999 to 2000, the rate was 13.72 percent. As a final check, we look at whether the rate of changing occupations—i.e., between private sector, public sector, and self-employment—changed after the reform. From 1997 to 1998, 1.76 percent of all employees switched from the public to the private sector, up from 1.70 percent between 1999 and 2000. During the same two time periods, 0.44 and 0.37 percent, respectively, switched from self-employment to the private sector.

Columns (5) and (6) of Table 1.4 cluster standard errors at the state $\times$ year (64 clusters) as well as at the industry $\times$ year (242 clusters) level to provide evidence on whether the group structure might be a serious issue in this setting. I 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 2.3). Clustering on the state $\times$ year level yields a standard error of 0.5992 and clustering on the industry $\times$ year level yields a standard error of 0.6857.

The last column of Table 1.4 tests whether there is evidence that the increase in statutory short-term sick pay had any effect on the incidence of long-term absenteeism. The estimated coefficient is almost zero in magnitude and not significant; thus it is reasonable to conclude that the distribution of long-term absence spells remained stable after the reform.

As has already been mentioned, an indirect method to test the common time trend assumption is to perform the same analyses for years with no reform. Significant reform estimates for years with no reform would cast doubts on the assumption of no unobserved year-group effects. In this context, however, this is not the case as Table 2.5 demonstrates.

Table 2.5: Difference-in-Differences Estimation on the Number of Absence Days: Placebo Estimates

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

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

2.5.2 Assessing Effect Heterogeneity and Further Reform Effects

Treatment Effect Heterogeneity and Health Effects

Table 2.6 displays extensive tests on treatment effect heterogeneity. Every column shows one OLS difference-in-differences model as in the main specification in column (1) of Table 2.3. The only difference is that the corresponding variable—with which we want to perform the heterogeneity test—is included both in levels and in interaction with the *DiD* regressor. Take as an example the first column of Panel A in the table. Here we want to check whether men have reacted differently from women to the reform. Hence, in addition to the gender dummy that was included in the model anyway, I interact the dummy variable *female* with *DiD* and run the model under the inclusion of this additional interaction term. The usual *DiD* point estimate tells us how men reacted to the reform. We find a highly significant 1.65 *DiD* estimate, suggesting that males reacted disproportionately to the reform. This is reinforced by the *DiD×female* point estimate. While the coefficient is imprecisely estimated, it is negative and large in magnitude. This provides some evidence that women did not react as strongly to the increase in sick leave benefits as men did, although the difference between men and women is imprecisely estimated.

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

Panel B exploits six (self-reported) health measures: self-assessed health (SAH), health satisfaction, a question on whether respondents feel impaired in their everyday tasks by their health status, and certified disability. Precisely how the six dummy variables are generated is explained in the notes to Table 2.6. The empirical results show that employees in bad health reacted much more strongly to the reform than the rest. Disabled employees, those with low health satisfaction, and those with low self-assessed health were induced by the reform to use between four and nine additional days of sick leave. This is a huge effect as compared to one day for the average population. By contrast, we find some evidence that healthy employees reacted less to the increase in insurance coverage than the average employee. There is even evidence that employees with very high health satisfaction did not react at all. This would be strong evidence against shirking—at least for this specific subgroup. Although the *DiD×high health satisfaction* coefficient in column (3) is only marginally significant, all three models that test the effects for healthy employees have negative signs on their interaction terms. Restricting the sample to respondents who indicated the best SAH category (equivalent to column (1)) and running the standard OLS-DiD model gives us an imprecisely estimated reform effect of 0.8. Using only respondents who were highly satisfied with their health yields an insignificant reform effect of 0.06 (equivalent to column (3)). Those who did not feel impaired by their health have an OLS-DiD coefficient of 1.01, which is significant at the five percent level, but still substantially lower than the one in column (1) of Table 2.3.

Panel C assesses effect heterogeneity with respect to objective workplace characteristics. Blue-collar workers seem to have reacted more strongly than white-collar workers. This might be because blue-collar workers work in less challenging jobs, which might lower their utility from work. Although not statistically significant, there are indications that employees of firms that had reduced their workforce between 1998 and 1999 took fewer days off than others, probably because they feared job loss. Interestingly, we find that the reform effect is negatively correlated with firm size. This adds to the credibility of my identification strategy since firm size is in general positively correlated with the use of sick leave, which can also be seen when looking at the firm size covariates in levels.[11] For the three firm size models, the triple interaction term always operates in the opposite direction of the simple firm size covariate. This is in line with our expectations, since we know that the bigger the firm, the more likely it is to have had a collective wage agreement and hence, the more likely it was that employees were not treated since the employer would have voluntarily provided 100 percent sick pay even before the reform. The same argument also holds for the last column in Panel C; the result reinforces my identification strategy once again: we find a significant 1.136 point estimate for employees who worked in a firm with no works council—in such firms it was very likely that sick pay did increase as a result of the reform. Again, the *no work council* covariate in levels works in the opposite direction.

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

[11] I always use *small firm* as reference category, which is why there is no *small firm* level covariate in column (4) of Panel C. When I use *very big firm* as reference category, the coefficient in levels for *small firm* is -3.8787 (std. err.: 0.5475).

Table 2.6: Assessing Heterogeneity in Reform Effects

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

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. All specifications estimate the model in equation (2.1) by OLS. Additionally, all models include an interaction term between *DiD* and the corresponding covariate in the column header. *Female* is a dummy variable with a one for females. *Over 40* is a dummy variable with a one for respondents over the age of 40. *Gross wage > median* is a dummy variable with a one for respondents who earn more than €2,199 per month. *Partner* has a one for respondents in a partnership. *Job loss previous year* is a dummy variable that indicates whether the employee changed the job in the previous year. *Highest school degree* means holding a certificate after 13 years of schooling. In Panel B, the first two columns make use of dummy variables that were generated from self-assessed health (SAH). *Health very good* has a one for respondents who indicated to have the best health status on the five-category SAH scale. *Health bad or poor* has a one for respondents who rated themselves in the worst two SAH categories. *Low health satisfaction* are the collapsed lowest four categories on an eleven-category scale on health satisfaction. *High health satisfaction* has a one for those ranked in the best health satisfaction category. *Not impaired by health* is generated from the answer category "Not at all" to the following question: "Aside from minor illnesses, does your health prevent you from completing everyday tasks like work around the house, paid work, studies, etc.? To what extent?" *Disabled* has a one for respondents who are officially certified as disabled. All models have 23,068 observations. The descriptive statistics for all column-header variables used are shown in Table 2.1.

Panel C: Objective workplace characteristics						
	white collar	**very big firm**	**medium size firm**	**small firm**	**workforce reduced btw. '98 & '99**	**no work council**
DiD × [column]	-1.3415*	-0.8821	0.0602	1.4605**	-1.1799	1.1356*
	(0.7259)	(0.8293)	(0.7831)	(0.7667)	(0.9686)	(0.6746)
DiD	2.0189**	1.5417**	1.3467*	1.1057	1.9689**	-0.6606
	(0.8423)	(0.7387)	(0.7412)	(0.7307)	(0.8025)	(0.7134)
Covariate [column]	0.1229	4.9069***	2.8659***		1.5741***	-1.5528***
	(0.8542)	(0.5967)	(0.5468)		(0.5511)	(0.5814)
Panel D: Subjective workplace characteristics						
	low job satisfaction	**very worried (job security)**	**job makes no fun**	**job loss likely in 2 yrs. ('98)**	**promotion likely in 2 yrs. ('98)**	
DiD × [column]	1.3064	-0.6709	0.2375	-0.8603	-0.4035	
	(2.4201)	(1.1887)	(2.061)	(1.8713)	(0.9446)	
DiD	1.2119*	1.5983**	1.7791*	1.9334**	1.8802**	
	(0.6796)	(0.7018)	(0.9991)	(0.7919)	(0.8332)	
Covariate [column]	4.0110***	1.2966*	1.1361	2.4224**	-0.4657	
	(1.2911)	(0.7185)	(0.9922)	(1.1416)	(0.5025)	

* 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 (2.1) by OLS. Additionally, all models include an interaction term between *DiD* and the corresponding covariate in the column header. The reference category of the *white collar* control variable in levels is blue collar worker. The reference category of the firm size controls (columns (2) to (4)) is always establishments with less than 20 employees (*small firm*). The covariates used in columns (1) to (4) of Panel C were sampled in all years. Thus, these models contain 23,068 observations each. The covariate used in column (5) of Panel C was solely surveyed in 1999 and has a one for respondents who claimed that their firm reduced the workforce in the year prior to the interview. The model has 17,201 observations. *No work council* in column (6) has a one for respondents working in firms without a work council. The question was only asked in 2001 and the model has 16,161 observations. *Low job satisfaction* stands for the lowest four categories on an eleven-category scale on job satisfaction (22,347 obs.) and *very worried about job security* has a one for respondents who answered "very concerned" to the following question: "Are you concerned about your job security" (22,503 obs.). *Job makes no fun* was only sampled in 1997 and has a one for those who answered "applies completely" or "applies more or less" towards the statement "I do not enjoy my work." (12,783 obs.). In 1998, respondents were asked whether they believe that they would loose their job (get promoted) within the next two years. Those who answered "very likely" or "likely" are represented by the dummy variables that are used in columns (4) and (5) of Panel D (16,771). For those variables that were only sampled in one specific year and that relate to the workplace, I 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 that do not relate to the workplace, I keep the respondents in all years in which they are in the sample. In both cases, time persistence is assumed. For example, respondents who indicated in 2001 that no work council exists at their workplace are kept in all years in which they had the same workplace as in 2001. I then assume the absence of a work council in all other years. The descriptive statistics for all column header variables used are shown in Table 2.1.

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

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. All specifications estimate the model in equation (2.1) by OLS. Additionally, all models include an interaction term between *DiD* and the corresponding covariate in the column header. The variables used in columns (1) to (3) were only sampled in 1997 (15,337 obs.). The variables used in columns (4) to (6) were only sampled in 1999 (17,351 obs.). Only respondents who answered the questions are kept in these models but they are kept in every sample year in which they answered the SOEP questionnaire. It is assumed that attitudes and personality traits remained stable over time. *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 2.1.

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

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

Table 2.7: Reform Effect on Employees' Health Status and Employers' Behavior

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

* $p<0.1$, ** $p<0.05$, *** $p<0.01$; standard errors in parentheses are adjusted for clustering on person identifiers. All specifications estimate the model in equation (2.1) by OLS but use the corresponding variable in the column header as outcome measure. All outcome variables used are detailed in Table 2.1. The model in the last column estimates the effect on gross wages for employees who claimed to have changed their jobs in the year prior to the interview (3,759 obs.). The models in the first three columns have 23,058 observations, the model in column (4) has 22,347 observations, and the models in columns (5) to (7) have 20,732 observations.

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

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

Labor Cost Effects and Employers' Reaction

Reform Induced Increase in Labor Costs

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

First, I assess how the increased obligation to provide sick leave benefits might have affected labor costs directly and indirectly. For the moment, we assume the world to be static. Then, the maximum overall increase in labor costs can easily be calculated by comparing the total employer-provided sick pay in the pre-reform years 1997/1998 with the total benefits in the post-reform years 1999/2000 under the assumption that every employer only provided the statutory 80 percent sick pay in the pre-reform years.[12] Thus, I calculate annual sick leave benefits for every employee in the sample and apply frequency weights to the sum. For the pre-reform period, I assume a replacement level of 80 percent of foregone gross wages and for the post-

[12] For this overall calculation, I do not need any of the regression results. This is a simple descriptive exercise, in which I make use of the full sample, i.e., I 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, I set the value for total absence days to 42, as only the first six weeks of sick leave are paid by the employer.

reform periods, I assume a replacement level of 100 percent. The frequency-weighted benefit sums for both periods are multiplied by the frequency-weighted number of employees in the treatment group. By taking the difference between pre- and post-reform years, we obtain a total maximum increase in labor costs of €5.153 billion for the two post-reform years.

This total increase in labor costs can be decomposed into three components. The first component is the intramarginal effect associated with the increase of the statutory sick pay level for the first six weeks from 80 to 100 percent of foregone gross wages. I approximate this amount by comparing the total sick leave payments in the pre-reform period to hypothetical sick leave payments for the same period and the same individuals, assuming that the sick pay was already increased to 100 percent at that time. I thus disentangle the direct labor cost effect from the effect that is induced by increasing absence rates as a consequence of the reform. Again, I do not need any regression results for this exercise and use the full sample. My calculation yields a direct labor cost effect of €3.87 billion for both years. If we assume that half of all firms had already provided 100 percent sick pay before the reform, this direct effect reduces to €1.93 billion.[13]

The second component represents the indirect labor cost effect, which was triggered by the reform-induced increase in workplace absence. From Table 2.3, we infer that the overall reform-induced increase in absence days equals approximately one day. Hence, I take the average daily gross wage in the pre-reform years and multiply it by the frequency-weighted number of employees in these years, resulting in an indirect labor cost effect of €1.61 billion. If we assume that the increase was 0.9 or 1.1 days, we get indirect effects of €1.45 and 1.77 billion over the two years, respectively.[14] The residual is the third component which is caused by time trends, changes in wages, and changes in the employment structure.

The total reform-induced increase in labor costs is thus $(1.93 + 1.61)/2 =$ €1.77 billion per year.[15]

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

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

[15] 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, I check the plausibility and sensitivity of this estimate. Using this method, I also control for panel attrition. To calculate the two effects, I multiply the official employment data by SOEP absence

I cross-check the plausibility of my 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. My 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.[16] Note that my estimate of €1.77 billion is net of time trends and assumes that 50 percent of all private-sector employees were actually treated. On the one hand, the similarity of my 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 my identification strategy and the assumption that about 50 percent of all private-sector employees were affected by the reform.

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

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

Since employers maximize profits, they must have responded in some way to the exogenous increase in labor costs of about €1.8 billion per year. In Germany at that time, very high total labor costs—especially in an international comparison—were a matter of serious concern for politicians, economists, and employers. These high labor costs were claimed to be the main barrier to job creation in Germany. Various researchers studied the relationship between labor costs and job losses by means of general macroeconomic equilibrium models (Zika, 1997; Feil et al., 2008; Meinhardt and Zwiener, 2005). If we relate the estimated increase in labor costs to the findings of these studies simply using the rule of proportion, we would obtain reform-induced job losses in the range of 40,000 to 80,000.[17]

rates and income data and get a similar estimate of $(2.21+1.98)/2 =$ €2.1 billion per year (German Federal Statistical Office, 1996, 1998).

[16] 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.

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

However, in Germany, dismissal protection is among the strictest worldwide. The very inflexible German labor market might have triggered other attempts at compensation as well. I provide empirical evidence on how employers might have reacted to the shock to labor costs in the last four columns of Table 7. Again, I use the same OLS-DiD model as before but now I use four different outcome measures. Column (5) measures the job turnover or mobility rate. The outcome measure indicates whether respondents changed jobs between the beginning of the year prior to the interview and the interview. We find the coefficient to be insignificant and very close to zero in magnitude.

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

In this context, it is important to know that, in Germany, there is strong tradition of autonomy in collective bargaining, which is also referred to as "Bismarckian corporatism." This means that the wage level and most other work conditions such as overtime compensation or fringe benefits are solely subject to negotiations between unions and employer's representatives. Politicians usually do not implement laws that target these fields.[18] While I do not claim that the relative increase in overtime and the relative decrease in wages can be unambiguously traced back to the increase in absence rates and labor costs, I argue that it is at least highly likely that substantial parts of these effects were triggered by the reform. As such, I have provided empirical evidence on how work conditions in a highly regulated labor market might be adjusted in equilibrium as a reaction to an increased obligation for employers to provide social insurance benefits.

reform.

[18] I 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.

2.6 Conclusion

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

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

The first part of the chapter showed, based on rich SOEP panel data, how increasing insurance coverage causally affected the sick leave behavior of employees. My identification strategy made use of conventional parametric difference-in-differences models but also non-parametric and combined approaches to prove the robustness of the results. Moreover, the panel data structure allowed us to eliminate or avoid the typical pitfalls of evaluation studies, such as selection effects and sample attrition. My findings suggest that the increase in statutory sick pay has led to an increase in sick leave of about one day per year and employee. This represents an increase of about 10 percent.

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

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

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

Chapter 3

Long-Term Absenteeism and Moral Hazard–Evidence from a Natural Experiment

Abstract

I theoretically and empirically disentangle the effects of cuts in the statutory sick pay levels on long-term absenteeism in Germany. The reforms have not induced significant changes in the average incidence rate and duration of sick leave periods longer than six weeks. The finding is theoretically confirmed assuming that the long-term sick are seriously sick. Thus, moral hazard seems to be less of an issue in the upper end of the sickness spell distribution. However, I find heterogeneity in the effects and significant duration decreases for certain subsamples. Finally, I calculate that within ten years, the cut in statutory long-term sick pay redistributed five billion Euros from the long-term sick to the insurance pool.

3.1 Introduction

The average number of sickness absence days per year and employee varies between 5 and 29 among the OCED countries (OECD, 2006). Average absence days are to a large degree determined by long-term absence spells. In Germany, which lies in the middle field of the ranking with 15 days, absence spells of more than six weeks account for 40 percent of all absence days although they only represent 4 percent of all sickness cases (Badura et al., 2008).

At the same time, legislative frameworks differ widely from one country to the next. In Europe, the statutory sickness absence insurance is integral part of the social insurance system. Typically, employers are obliged to provide sick pay for short-term absences, whereas health insurance providers or taxpayers compensate wage losses for the long-term sick. The U.S. do not know a statutory sickness insurance for short-term absentees on the federal level. However, the U.S. and Canada know the *workers' compensation insurance (WCI)* that is administered on a state-by-state basis and covers incomes losses due to *work-related* sickness or injury. On the federal level, the *disability insurance (DI)* replaces income losses stemming from a permanent labor market withdrawal due to work disability.

The literature on sickness absence in general is quite rich. It has been found that workplace conditions determine sick leave behavior (Dionne and Dostie, 2007) as well as probation and work contract periods(Engellandt and Riphahn, 2005; Ichino and Riphahn, 2005), the level of employment protection (Riphahn, 2004), and economic upswings or downturns (Askildsen et al., 2005). However, empirical evidence concerning the relationship between the design of the sickness insurance scheme and sick leave behavior is scanty at best, especially as compared to other fields like the vast literature on unemployment benefits and unemployment duration. Some studies from Sweden have shown that employees adapt their sick leave behavior to changes in replacement levels (Johansson and Palme, 2002; Henrekson and Persson, 2004; Johansson and Palme, 2005). Moreover, Puhani and Sonderhof (2010) have shown that changes in statutory short-term sick pay affected the sick leave behavior in Germany. All studies cited above explicitly analyze the effects on short-term sickness absences within the European statutory sickness insurance. There is also empirical evidence from North America on the workers' compensation insurance (WCI) and the disability insurance (DI), althought the findings are inconclusive. While Meyer et al. (1995) find that an increase in WCI benefits in 1987 has led to increased injury duration, the results from the Curington (1994) study using data from the 1960s and 1970s are mixed. Besides the WCI, the DI has attracted a lot of attention among economists. Many studies find that the generosity of the DI affects labor supply behavior on the extensive margin (Bound, 1989; Gruber, 2000; Chen and van der Klaauw, 2008), although there is also convincing evidence that this might not be the case (Campolieti, 2004). Researchers have also studied the DI application process (e.g. de Jong et al. (2010)).

It is very important to keep in mind that the empirical findings concerning the DI and the WCI are unlikely to be directly transferable to the sickness absence insurance. While the European statutory sickness insurance covers all types of work-related *and* work-unrelated illnesses, the WCI solely covers the special case of a work-related illness or injury. On the other hand, both social insurances have in common that employees are still employed while being on sick leave. Thus, both focus on labor supply behavior on the *intensive* margin. In contrast to that, the DI deals with labor supply behavior on the *extensive* margin and hence a complete withdrawal from the labor market.

To the best of my knowledge, this is the first study that explicitly analyzes the impact of cuts in statutory long-term sick pay on long-term absenteeism, i.e., on sickness spells of more than six weeks. In Germany, statutory long-term sick pay is provided by the Statutory Health Insurance (SHI) system. In 1996, the total benefit sum amounted to €9.3 billion, comprising 7.3 percent of all expenditures by the SHI system. At that time, two health reforms were implemented, both of which cut the level of paid sick leave. I theoretically and empirically analyze the effects of both reforms on long-term absenteeism. Additionally, I calculate the reform-induced SHI savings and redistributional effects.

In the remainder of this manuscript, sickness spells that last less than six weeks are defined as short-term absenteeism and sickness spells that last longer than six weeks are defined as long-term absenteeism. Analogously, statutory sick pay during the first six weeks of a spell is defined as statutory short-term sick pay, while statutory long-term sick pay refers to episodes of more than six weeks.

The first reform cut statutory short-term sick pay from 100 to 80 percent of foregone gross wages, whereas the second reform cut statutory long-term sick pay from 80 to 70 percent of foregone gross wages. Both reforms generate exogenous sources of variation and yield testable implications.

To theoretically predict the effects of both reforms on long-term absenteeism, I employ a simple dynamic model of absence behavior. First, if moral hazard plays a role and employees on long-term sick leave react to economic incentives, the cut in long-term sick pay should lead to a decrease in long-term absenteeism as the direct costs of being on long-term sick leave unambiguously increase. However, short-term sick pay was likewise cut at the same time. Since the cut in short-term sick pay was stronger than the cut in long-term sick pay, this might have triggered an indirect effect. Hence, second, from a theoretical point of view, the two reforms jointly may have affected long-term absenteeism in a positive way since the costs of long-term absences decreased relative to the costs of short-term absences. In other words, the gap in the replacement levels between short-term and long-term sick leave decreased as a consequence of the reforms. Later on, in Section 3.3, I derive the direct and the indirect effect by means of a simple theoretical model. However, under the assumption that employees on long-term sick leave are seriously sick, the incentive structure of the sick pay scheme would break down and individuals would not adapt their labor supply behavior to moderate cuts in sick pay.

Since Germany has two independent health care systems existing side by side, I am able to identify subsamples that were affected by none, one, or both of the reforms (see next section). Then, using data from the German Socio-Economic Panel Study (SOEP) and difference-in-differences methods, I can estimate the net and the direct effect of the two reforms on the incidence and duration of long-term absence spells.

My empirical findings in Section 3.6 indicate that, on population average, the cut in replacement levels did not affect the incidence and duration of long-term sickness spells, either directly or indirectly. This result is in line with my model predictions under the assumption that employees on long-term sick leave are indeed seriously sick. However, I find evidence of heterogeneity in the effects. For the poor as well as for middle-aged persons employed full-time, the duration of long-term absenteeism decreased significantly. Overall, my findings suggest that employees who have been on certified sick leave for more than six weeks are not very responsive to moderate monetary incentives, which implies that, in contrast to shorter absence spells, moral hazard is of less importance in the upper end of the sickness spell distribution. In the last subsection before I conclude, I calculate that from 1997 to 2006, the cut in statutory long-term sick pay redistributed five billion Euros from the long-term sick to the statutory health insurance pool for the benefit of lower contribution rates.

3.2 The German Health Care System and the Policy Reforms

3.2.1 The Two Track German Health Care System

The German health care system actually consists of two independent health care systems existing side by side. The more important of the two is the Statutory Health Insurance (SHI) system, which covers about 90 percent of the German population. Employees whose income from salary is below a politically defined income threshold (2007: €3,975 per month) are compulsorily insured under the SHI. High-income earners who exceed that threshold, as well as the self-employed, have the right to choose between the SHI or a private health insurance provider. Non-working spouses and dependent children of individuals insured under the SHI are automatically insured by the SHI family insurance at no charge. Special groups such as students or unemployed are subject to special arrangements but are mostly insured under the SHI. In principle, insurance coverage is the same for all those insured under the SHI (German Ministry of Health, 2008).

The second component of the German health care system is Private Health Insurance (PHI). It basically covers private-sector employees who earn above the income threshold, public sector employees, and self-employed persons. Privately insured people pay risk-related insurance premiums determined by an initial health checkup. The premiums exceed the expected expenditures in younger age brackets, since health insurance providers build up reserves for rising expenditures with increased age. Coverage is provided under a range of different health plans, and insurance contracts are subject to private law. Consequently, in Germany, public health care reforms apply only to the SHI, not to the PHI.

It is important to keep in mind that compulsorily insured persons have no right to choose the health insurance system or benefit package. They are compulsorily insured under the standard SHI insurance scheme. Once an optionally insured person (a high-income earner, self-employed person, or civil servant) opts out of the SHI system, it is practically impossible to switch back into it. Employees above the income threshold are legally forbidden from switching back, while employees who fall below the income threshold in subsequent years may do so under certain conditions, but are not able to carry along the reserves that their PHI providers have built up since these are not portable (neither between PHI and SHI, nor between the different private health insurance providers).[1] In reality, switching to a private health insurance provider may be regarded as a lifetime decision, and switching between the SHI system and PHI – as well as between PHI providers – is therefore very rare.

3.2.2 The German Statutory Sick Pay Scheme

If an employee falls sick, a certificate from a physician is required from the fourth day of sick leave. The employer is legally obliged to provide statutory short-term sick pay up to six weeks per sickness spell regardless of the employee's health insurance. From the seventh week onwards, the physician needs to issue different certificates at reasonable time intervals of usually one week, and long-term sick pay is provided by the SHI or the PHI. The replacement level for persons on long-term sick leave insured under the SHI is codified in the social legislation and is the same for all those with SHI insurance. In 1996, SHI payments for long-term absenteeism made up 7.3 percent of all SHI expenditures, which equaled 9.3 billion euros (German Federal Statistical Office, 1998). Employees insured under the PHI insure the risk of falling long-term sick individually.

The system for monitoring employees on sick leave is a potentially important determinant of the degree of moral hazard in the insurance market. In Germany, the *Medizinischer Dienst der Gesetzlichen Krankenversicherungen* (Medical Service of the SHI) exists for this purpose. One of the original objectives of the medical service was to monitor absenteeism. It is explicitly stated in the guidelines of this institution that long-term absenteeism in particular should be prevented in order to reduce the risk of patients descending the social ladder (Medizinischer Dienst der Krankenversicherung (MDK), 2008). The German social legislation stipulates that the SHI is obligated to call upon the Medical Service to provide an expert opinion, in order to dispel any doubts about work absences. Such doubts may arise if the insured person is absent unusually often or repeatedly sick for short-term periods on Mondays or Fridays. If physicians certify sickness uncommonly often, the SHI may ask for an expert opinion. The employer also has the right to call upon the Medical Service to provide an expert opinion. Expert opinions are based on available

[1] Until 2009, accrued reserves for rising health expenditures with increased age were not portable at all. From January 1, 2009 on, portability of accrued reserves between PHI providers has been made compulsive to a strictly defined extent.

medical documents, information about the workplace, and a compulsory statement from the patient. If necessary, the medical service has the right to examine the patient physically and to cut benefits.[2] In 2007, about 2,000 full-time equivalent employees and independent physicians worked for the medical service and examined 1,719,386 cases of absenteeism (Medizinischer Dienst der Krankenversicherung (MDK), 2008).

3.2.3 The Policy Reforms

Two health reforms were implemented at the end of 1996. First, from October 1996 on, the replacement level during the first six weeks of a sickness episode (i.e., statutory short-term sick pay) was reduced from 100 to 80 percent of foregone gross wages.[3] This reform had, at least theoretically, an indirect impact on sickness spells of more than six weeks and should therefore be considered. Second, the replacement level for absence spells of more than six weeks (i.e., statutory long-term sick pay) was cut from 80 to 70 percent of foregone gross wages for those insured under the SHI. This health reform act became effective on January 1, 1997. [4] Figure 3.1 illustrates how the two reform worked.

SHI statutory long-term sick pay is additionally limited by two benefit caps. First, if the wage of an employee insured under the SHI exceeds the legally defined contribution ceiling, then long-term sick pay is limited to 70 (80) percent of this contribution ceiling (2009: 0.7*€3,675 per month) as contributions are capped over this ceiling as well. Second, before 1997, the replacement level was 80 percent of the gross wage if the total amount did not exceed 100 percent of the net wage after taxes and social contributions. For example, a worker might earn €2,500 gross per month and €1,800 net per month. Then, the basic rule implied statutory long-term sick leave that amounted to 0.8*€2,500=€2,000. However, the second benefit cap

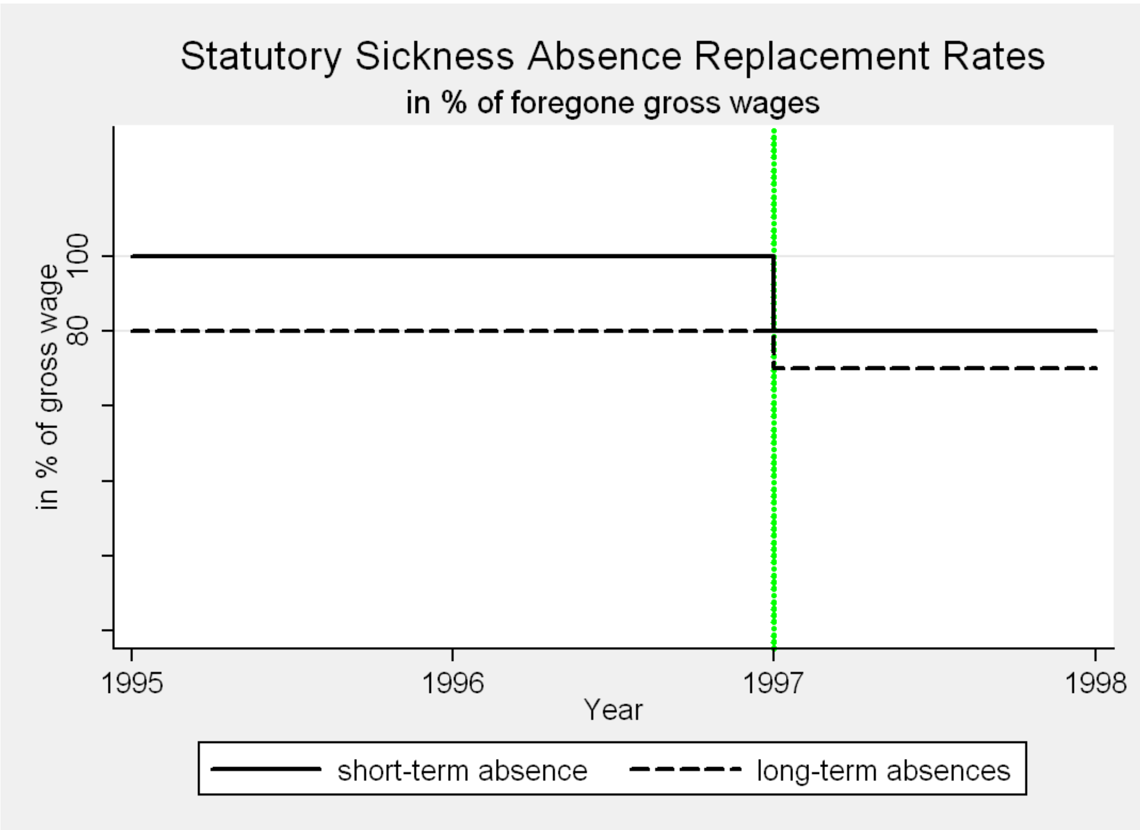
[2] The wordings of the law can be found in the Social Code Book V, article 275, para. 1, 1a; article 276, para. 5.

[3] Passed on September 15, 1996 this law is the *Arbeitsrechtliches Gesetz zur Förderung von Wachstum und Beschäftigung (Arbeitsrechtliches Beschäftigungsförderungsgesetz), BGBl. I 1996 p. 1476-1479.* It became effective at October 1, 1996. It should be noted that I am not able to precisely identify those employees who were effectively affected by this law, as employers and unions voluntarily agreed in some collective wage agreements to continue the old sick pay scheme. However, in principle, the law applied to all private sector employees whom I define below as being treated by the cut in statutory short-term sick pay. Using all private sector employees jointly as treatment group, Ziebarth (2009) have shown that the cut in statutory short-term sick pay reduced short-term absenteeism.

[4] Passed on November 1, 1996, this law is the *Gesetz zur Entlastung der Beiträge in der gesetzlichen Krankenversicherung (Beitragsentlastungsgesetz - BeitrEntlG), BGBl. I 1996 p. 1631-1633.*

limited the benefit to €1,800 per month before the reform. The cut in statutory long-term sick pay decreased the replacement level to 70 percent of the gross wage (i.e., 0.7*€2,500=€1,750) and the benefit cap to 90 percent of the net wage (i.e., 0.9*€1,800=€1,620). As can be seen by means of this little example, benefit caps were also decreased in the course of the reform, depending on the relation between gross and net wages – which in turn is determined by the income level, the marital status, and the number of children – employees insured under the SHI were affected differently by the cut in long-term sick pay. This introduces additional exogenous variation which allows me to generate an index that mirrors the cut in statutory long-term sick pay for each individual on a continuous scale from zero percent of the gross wage up to 10 percent of the gross wage.

Figure 3.1: Replacement Levels for Short and Long-Term Absence Spells



Independent from the reforms analyzed in this chapter, the German sick pay scheme exerts an incentive to substitute a long-term spell by several short-term spells since statutory sick pay for the latter is higher. However, German social legislation explicitly forbids such substitution of spells: If employees repeatedly call-in sick due to the same illness, they are no longer entitled to employer-provided statutory short-term sick pay.[5]

I now define subsamples that have been affected differently by the two health reforms, thereby serving as treatment and control groups in the evaluation of this natural experiment. As the sickness compensation for long-term absence is paid for by the health insurance and not by the employer, the second reform did not affect privately insured people, whose long-term sick leave replacement levels are subject to individual insurance contracts.

Table 3.1: Definition of Subsamples

	Cut statutory short-term sick pay (employer)	**Cut statutory long-term sick pay (SHI)**
Private sector employees with SHI (1)	yes	yes
(*Treatment Group 1*)		
Public sector employees with SHI (2)	no	yes
Trainees with SHI (3)	no	yes
(*Treatment Group 2*)		
Public sector employees with PHI (4)	no	no
Self-employed with PHI (5)	no	no
(*Control Group*)		

Table 3.1 shows that private-sector employees who were insured with the SHI (subsample (1)) were affected by both reforms. In contrast, SHI-insured public-sector employees (subsample (2)) were affected by the cut in statutory long-term sick pay but not by the cut in statutory short-term sick pay due to political decisions. The same holds for SHI-insured trainees (subsample (3)). While subsample (1) is defined as *Treatment Group 1*, subsamples (2) and (3) are called *Treatment Group*

[5] *Gesetz über die Zahlung des Arbeitsentgelts an Feiertagen und im Krankheitsfall (Entgeltfortzahlungsgesetz - EntgFG), BGBl. I 1994 p. 1014, 1065.* Para. 3 contains the passage.

2. The last two subsamples, PHI-insured public-sector employees and PHI-insured self-employed persons, were not affected by any of the reforms and are called *Control Group*.

3.3 A Dynamic Model of Absence Behavior

In the following, I analyze the absence behavior of an individual i within a two-period model. I modify a model by Brown (1994) so as to be able to study the theoretical effects of the German health reforms on long-term absence behavior. The individual's utility function can be specified as:

$$u_t = (1 - \sigma_t)c_t + \sigma_t l_t, \qquad t = t, t + 1;\ \sigma_t \in [0, 1] \tag{3.1}$$

where t is the time period, c_t represents consumption in period t, and l_t leisure in period t. The sickness level in t is specified by σ_t, where larger values of σ_t represent a higher degree of sickness. If the sickness index tends towards unity, i.e., a high level of sickness prevails, the individual draws utility only from leisure or recuperation time rather than consumption. On the other hand, if the sickness level is relatively low, the individual attaches more weight to consumption as opposed to leisure. To simplify the analysis, I assume that $f(\sigma_t)$ follows a uniform distribution:

$$f(\sigma_t) = \begin{cases} 1 \quad if \quad 0 \leq \sigma_t \leq 1 \\ 0 \qquad otherwise \end{cases}$$

Hence each sickness level is equally probable. At time t, individuals are aware of their sickness level σ_t but concerning the subsequent period, only the probability distribution $f(\sigma_{t+1})$ is known.

To adequately model the German sick pay scheme, I define the statutory long-term sick pay as r_l with $0 < r_l < 1$ and the statutory short-term sick pay as r_h with $0 < r_h < 1$. Moreover, $r_l < r_h < w$, where w represents the gross wage and is normalized to one. Sick pay is always provided when the individual is absent from work. Long-term sickness is when an individual is on sick leave for at least two continuous periods. Hence, in the first absence period after a working period, the sick pay is r_h, which is reduced to r_l in the second period. If a working period follows a long-term sickness period, the replacement level for the next sickness period is again r_h.

A key feature of this simple dynamic model is the concept of the reservation sickness level, σ_t^*, as introduced by Barmby et al. (1994). The reservation sickness level is defined as the value of σ_t such that an individual is indifferent between going to work and staying home. To be more precise, at σ_t^* the utility from working in period 1 plus the expected utility in period 2 equals the utility from being absent in period 1 plus the expected utility in period 2. As I am primarily interested in the reform effects on long-term absenteeism, I assume that the individual was on sick leave in $t-1$ and is eligible for sick pay in t with r_l as the replacement level. In t, the reservation level is hence implicitly defined by:

$$(1-\sigma_t^*)r_l + \sigma_t^* T + \frac{1}{1+\rho}E(U_{t+1}^{absent}) = (1-\sigma_t^*)w + \sigma_t^*(T-h) + \frac{1}{1+\rho}E(U_{t+1}^{work}) \tag{3.2}$$

The left hand side of this equation represents the utility in period t if the individual continues to be on sick leave with sick leave compensation r_l and leisure T, where T is the total time available. The expected utility from period $t+1$ is added and discounted with the individual's time preference rate ρ. Analogously, the right hand side adds up the discounted utility in $t+1$ with the utility from working h hours and enjoying $T-h$ hours leisure in t.[6]

The individual decides whether to be absent from work by maximizing utility over both periods. If $\sigma_t > \sigma_t^*$, i.e., the actual sickness level exceeds the reservation sickness level, the individual stays away from work as more weight is placed on leisure rather than consumption. In other words, if employees are seriously sick, they value recuperation time far more than materialistic needs and go on sick leave. On the other hand, if $\sigma_t < \sigma_t^*$, individuals maximize their utility by working h hours.

One has to bear in mind that the decision to be absent from work or not has implications for the sick pay level in the next period. If individuals are absent from work in t, they get r_l in t as well as in $t+1$ – if their sickness continues to be so severe that $\sigma_{t+1} > \sigma_{t+1}^{a*}$, where σ_{t+1}^{a*} is the reservation sickness level in $t+1$ conditional on having been absent in t. If they work in t and fall sick in $t+1$, with $\sigma_{t+1} > \sigma_{t+1}^{w*}$, their sick pay is r_h. Hence I can define $E(U_{t+1}^{absent})$ which is the expected utility in

[6] I assume a rigid employment contract without the possibility of working overtime or less than the contracted hours h.

$t+1$ conditional on having been absent at time t:

$$\begin{aligned} E(U_{t+1}^{absent}) \quad = \quad & (1-\sigma_{t+1}^{a*})\left[\left(1-\left(\frac{1+\sigma_{t+1}^{a*}}{2}\right)\right)r_l+\left(\frac{1+\sigma_{t+1}^{a*}}{2}\right)T\right]+ \\ & \sigma_{t+1}^{a*}\left[\left(1-\left(\frac{\sigma_{t+1}^{a*}}{2}\right)\right)w+\left(\frac{\sigma_{t+1}^{a*}}{2}\right)(T-h)\right] \end{aligned} \tag{3.3}$$

As can be seen from (3.3), the expected utility in $t+1$ is expressed as the weighted average of the expected utility from attending work and being absent from work. The weights represent the probability that σ_{t+1} is less than the reservation sickness level and exceed the reservation sickness level, respectively. The expected values of consumption and leisure are evaluated by using the conditional probability distribution. Conditional on σ_{t+1} being between 0 and σ_{t+1}^{a*}, the expected value of σ_{t+1}, which is $\frac{\sigma_{t+1}^{a*}}{2}$ for the uniform distribution, is taken to evaluate the utility of a working employee. Analogously, the expected value of σ_{t+1}, conditional on being between σ_{t+1}^{a*} and 1,$\frac{1+\sigma_{t+1}^{a*}}{2}$, is substituted into the utility function for an absent employee.

Equivalently defined is $E(U_{t+1}^{work})$ which is the expected utility in $t+1$ conditional on having worked in t:

$$\begin{aligned} E(U_{t+1}^{work}) \quad = \quad & \sigma_{t+1}^{w*}\left[\left(1-\left(\frac{\sigma_{t+1}^{w*}}{2}\right)\right)w+\left(\frac{\sigma_{t+1}^{w*}}{2}\right)(T-h)\right]+ \\ & (1-\sigma_{t+1}^{w*})\left[\left(1-\left(\frac{1+\sigma_{t+1}^{w*}}{2}\right)\right)r_h+\left(\frac{1+\sigma_{t+1}^{w*}}{2}\right)T\right] \end{aligned} \tag{3.4}$$

Finally, I derive σ_{t+1}^{a*} and σ_{t+1}^{w*} as:

$$\sigma_{t+1}^{a*}=\frac{w-r_l}{w-r_l+h} \tag{3.5}$$

$$\sigma_{t+1}^{w*}=\frac{w-r_h}{w-r_h+h} \tag{3.6}$$

We find that $\frac{\partial\sigma_{t+1}^{a*}}{\partial r_l}<0$ and $\frac{\partial\sigma_{t+1}^{w*}}{\partial r_h}<0$, which means that a decrease in sick pay levels has a positive impact on the reservation sickness levels, resulting, ceteris paribus, in a lower probability to be absent from work. This is what we would expect intuitively when the costs of sickness rise. Moreover, static labor supply models also predict a decrease in absenteeism with decreasing sick pay rates (Brown and Sessions, 1996). Henceforth, I call this the direct effect of a reduction in sick pay.

As $r_l < r_h < w$, we get $\sigma_{t+1}^{a*} > \sigma_{t+1}^{w*}$ meaning that the probability to work in $t+1$ is higher for an employee who stayed home in t as opposed to an employee who worked in t. The reason is that the gap between wages and sick pay, i.e., the cost of absence, is bigger for long-term absenteeism as compared to a short-term absenteeism. This is a reasonable approximation of the statutory sick leave regulations in Germany.

Plugging equations (3) to (6) into (2) and solving for the reservation sickness level σ_t^* yields:

$$\sigma_t^* = \sigma_{t+1}^{a*} + \frac{\varpi}{(1+\rho)(w - r_l + h)} \tag{3.7}$$

$$\varpi = \frac{(r_h - r_l)h^2}{2(w - r_l + h)(w - r_h + h)} > 0 \tag{3.8}$$

We see that σ_t^* equals σ_{t+1}^{a*} plus a discounted positive term which I interpret as the impact of future absence costs on the today's decision to be absent from work or not. It illustrates how the German sick pay scheme, which penalizes long absence spells more severely than short absence spells, impacts the probability to stay at home in the current period. In the case of a flat sick pay level, which would not depend on the length of absence, the second term would vanish and the probability of being absent from work today would equal the probability of being absent from work tomorrow. Remember that this holds under the assumption that every health status is equally probable and outside the individual's influence. Utility-maximizing individuals need to take the impact of today's absence behavior on future sick pay entitlements into account.

I now predict how long-term absenteeism is affected if the sick pay levels for short and long absence spells decrease and the employee is entitled to r_l in case of being absent. Consider first the effects of a reduction in r_l.

$$\frac{\partial \sigma_t^*}{\partial r_l} = \underbrace{\frac{\partial \sigma_{t+1}^{a*}}{\partial r_l}}_{<0} + \underbrace{\frac{\frac{\partial \varpi}{\partial r_l}(w - r_l + h) + \varpi}{(1+\rho)(w - r_l + h)}}_{<0} \tag{3.9}$$

We see from equation (9) that the total effect of a decrease in r_l is the sum of the direct effect $\frac{\partial \sigma_{t+1}^{a*}}{\partial r_l}$ and an additional factor. Hence, it is crucial to consider the impact of the discounted future term when evaluating the impact of a reduction in r_l. The second term represents the indirect effect that arises from the gap in the replacement levels between long and short-term absence spells, $r_h - r_l$. In case of a flat compensation scheme the gap closes and the indirect effect disappears. Ceteris paribus, a reduction in r_l widens the compensation gap, increases future absence costs, and thus affects long-term absenteeism negatively, thereby strengthening the direct effect.

Now I consider a reduction in r_h. Note that there is no direct effect of a decrease in r_h for people in an ongoing long-term sickness spell. These people continue to get r_l if they remain absent, and get their full wage if they go back to work. However, a reduction in r_h would, ceteris paribus, diminish the compensation gap between short and long-term absences and thus exert a positive effect on long-term absenteeism.

$$\frac{\partial \sigma_t^*}{\partial r_h} = \underbrace{\frac{\partial \sigma_{t+1}^{a*}}{\partial r_h}}_{=0} + \underbrace{\frac{\frac{\partial \varpi}{\partial r_h}}{(1+\rho)(w - r_l + h)}}_{>0} \tag{3.10}$$

I now want to relax the rather restrictive assumption that the sickness level σ_t is independent of the sickness level in the previous period and that every sickness level is equally probable in every period. Suppose that the sickness levels are serially correlated and that r_h is paid for sickness spells up to six periods. If the employee continues to be on sick leave in the seventh period, r_l is paid. For a sickness spell to last more than six periods, the illness must to be so severe that $\sigma_t > \sigma_t^*$ in every period. If that is the case, the incentive structure of the sick leave scheme breaks down and the employee is absent from work in every period. Hence, if employees are seriously sick, which means that their degree of sickness tends towards unity, and the replacement levels change only moderately without taking on extreme values, then these employees do not react to economic incentives.

In Section 3.6, I empirically estimate the effects which I derived above theoretically.

3.4 Data and Variable Definitions

The data set that I use in the empirical part is the German Socio-Economic Panel Study (SOEP). The SOEP is an annual representative household survey that was started in 1984 and sampled more than 20,000 persons in 2006. Further details can be found elsewhere (Wagner et al., 2007).

For the core analyses, I use data of the years 1994 to 1999. As my goal is to evaluate a reduction in wage compensation levels, I drop non-working respondents and those who are not eligible for long-term sickness compensation (i.e., people who earn less than €400 per month and working students). Furthermore, I drop observations with item-non response and restrict the sample to respondents aged 18 to 65.

3.4.1 Dependent Variables and Covariates

The SOEP contains various questions about the usage of health services and health insurance. I generate my first dependent dummy variable, *longabs*, which measures the incidence of long-term absenteeism, from the following question that was asked continuously from 1994 on: *"Were you sick from work for more than six weeks at one time in 19XX?"* Since sick pay decreases after six weeks, since it is no longer disbursed by the employer but by the health insurance, and since a different certificate needs to be issued by the physician, measurement errors should play a minor role here.

To measure how many days long-term sick pay was received, I use the following SOEP question: *"How many days were you not able to work in 199X because of illness?"* I generate my second dependent variable by subtracting, for those who had a long-term absence spell, the number of employer-paid sick days – namely 30 for the first six weeks – from the total number of days absent. This variable is called *longabsdays* and measures the duration of long-term absenteeism.[7] [8] Clearly, this duration indicator is subject to measurement errors as I assume that the respondents had no other absence spells. Moreover, comparing the mean value of *longabsdays* with official data, it becomes clear that we face a systematic underreporting in the survey data, as persons with long-term sickness spells are less likely to participate

[7] Public sector employees enjoy special privileges. In contrast to private sector employees, they receive 100 percent sick pay up to 26 weeks depending on seniority. Since I have detailed information about the seniority levels, I am able to identify these privileged public sector employees. For them, I redefine long-term absence spells as sickness spells for which they receive the lower SHI statutory sick pay.

[8] For those respondents who indicated having been absent for more than six weeks but who reported a total number of sick days of less than 30, I replace the values on *longabsdays* with a one. By estimating a Zero-inflated Negbin-2 model and predicting the total number of benefit days, I impute missing values for respondents with item-non response on the variable about total sick leave days. I impute the values only for respondents who indicated that they were on long-term sick leave and who had no missings on the other covariates.

in the survey. However, as long as the cut in statutory long-term sick pay did not affect the probability to participate in the survey and did not affect the sickness spell distribution, this duration measure should be sufficient to assess the reform effects. While the former assumption is likely to hold, one could argue that the latter is more problematic. Those who were only affected by the cut in statutory long-term sick pay had an increased incentive to substitute long-term spells by short-term spells. However, according to German law, the eligibility for employer-provided statutory short-term sick pay expires in case of such sickness spell substitutions (see Section 3.2.3 for more details). Once more, the importance of having various treatment groups is emphasized here. By comparing *Treatment Group 1* with the *Control Group*, I cannot unambiguously identify reform effects on the duration of long-term workplace absences, since a negative effect on *longabsdays* might have been triggered by the cut in short *or* long-term sick pay. However, contrasting *Treatment Group 2*, which was affected only by the cut in long-term sick pay, with the *Control Group*, and bearing in mind that sickness spell substitutions are no issue in this setting, I can estimate the impact of the cut in statutory long-term sick pay on the length of long-term sickness spells.

Since both questions on absenteeism, and thus both dependent variables, refer to the last calendar year, I use information of time variant covariates from the previous year if the respondent was interviewed the year before. For respondents who were not interviewed in the previous year, I use the current values of their covariates and assume that they did not change since the onset of the long-term sick leave episode.

The whole set of explanatory variables can be found in Appendix B. It is categorized as follows: A first group of covariates incorporates variables on personal characteristics, like the dummies *female, immigrant, East Germany, partner, married, children, disabled, good health, bad health, 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 the *monthly gross wage*.

3.4.2 Treatment Indicators and Treatment Intensity Indices

As described in Section 3.2.3 and visualized in Table 3.1, I define different subsamples as *Treatment Group 1*, *Treatment Group 2*, and *Control Group*. Since the SOEP is very detailed about the insurance status and the workplace of the respondents, I can precisely assign respondents to the different groups. However, self-employed persons insured under the SHI have the option to opt out of long-term sick pay in order to obtain lower contribution rates. Since I am unable to identify respondents with such contracts, I drop them.

An can be seen in Table 3.2, I generate three treatment dummy indicators that I use below in my empirical models to estimate the direct, indirect, and net effect of the two sick pay reforms on long-term absenteeism. *T1* has a one for all employees who were affected by both reforms (*Treatment Group 1*) and a zero for all those who were affected by none (*Control Group*). I use *T1* to estimate the reforms' net effet on long-term absenteeism. To disentangle the direct effect, I employ *T2* which has a one for all respondents who were solely affected by the cut in statutory long-term sick pay (*Treatment Group 2*) and a zero for the *Control Group*. In contrast, *T3* has a one for *Treatment Group 1* and a zero for *Treatment Group 2*, helping me in assessing the indirect effect.

Table 3.2: Definition of Treatment Indicators to Estimate Reform Effects

Effect to be estimated	**Treatment Indicator**	**=1**	**=0**
Net effect	T1	subsample (1) (*Treatment Group 1*)	subsamples (4) + (5) (*Control Group*)
Direct effect	T2	subsamples (2) + (3) (*Treatment Group 2*)	subsamples (4) + (5) (*Control Group*)
Indirect effect	T3	subsample (1) (*Treatment Group 1*)	subsamples (2) + (3) (*Treatment Group 2*)

As discussed in Section 3.2.3, not only was statutory long-term sick pay cut from 80 to 70 percent of foregone gross wages but likewise was its benefit cap decreased from 100 to 90 percent of the net wage after taxes and social contributions. Depending on the relation between gross and net wage, this reform element generated an additional source of exogenous variation in terms of treatment intensity. As a result, individuals experienced cuts in their statutory long-term sick pay from zero up to ten percent of their gross wage. Thus, I calculate for each individual his or her (potential) reform induced decrease in statutory long-term sick pay relative to the gross wage. This is feasible since the SOEP samples data on gross wages, net wages, and other income components such as Christmas or vacation bonuses. The SOEP group deals precisely with the problem of item-non response and imputes missing values thoroughly (Frick and Grabka, 2005).

Then, in addition to the three treatment dummy indicators, I generate two continuous treatment intensity indices. Both sample the same individuals as *T1* and *T2* and are called *T1index* and *T2index*. *T1index* has the value 0 for those in the *Control Group* and values from 0.57 up to 10.00 for those in *Treatment Group 1*, meaning that the decrease in statutory long-term sick pay varied between 0.57 and 10 percent of the employees' gross wage. Equivalently built is *T2index*. Everyone in the *Control Group* has a zero on *T2index* and employees in *Treatment Group 2* have positive values up to 10.00. The density of *T1index* and *T2index* peaks around six and ten. About 80 percent of the treated faced a cut in statutory long-term sick pay between 4 and 8 percent of their gross wage and about 12 percent experienced a cut of 10 percent of their gross wage.

3.5 Estimation Strategy and Identification

3.5.1 Probit Specification

To estimate the causal reform effects on the incidence of long-term absence spells, I fit a difference-in-differences (DiD) probit model of the following type:

$$Pr(y_{it} = 1) = \Phi(\alpha + \beta p97_t + \gamma D_{it} + \delta DiD_{it} + s'_{it}\psi + \rho_t + \phi_s) \tag{3.11}$$

where y_{it} stands for the incidence of long-term absenteeism, *longabs*, for individual i in year t. The dummy $p97_t$ has a one for post-reform years and a zero for pre-reform years. Depending of the empirical specification, the treatment variable D_{it} stands representative for *T1*, *T2*, *T3*, *T1index*, or *T2index* (see Section 3.2.3 and Section 3.4.2). My variable of interest, DiD_{it}, can be interpreted as the interaction term between D_{it} and $p97_t$ and takes on positive values for treated individuals in post-reform years. By including time dummies ρ_t I control for common time shocks that might affect long-term absenteeism. State dummies ϕ_s account for permanent differences across the 16 German states along with the annual state unemployment rate that controls for changes in the tightness of the regional labor market and that is included in the $K \times 1$ column vector s'_{it}. The other $K-1$ regressors are made up of personal controls including health status, educational controls, and job-related controls as shown in Appendix B.

Should the assumptions discussed below hold, the marginal effect of the interaction term DiD_{it} gives us the causal reform effect and is henceforth always displayed when output tables are presented.[9]

3.5.2 Count Data Specification

To estimate how the policy reforms affected the duration of long-term absence spells in post-reform periods, I fit count data models. Since the second dependent variable *longdaysabs* is a count with excess zero observations and overdispersion, i.e., the conditional variance exceeding the conditional mean, count data models should capture these distributional properties appropriately. Based on the Akaike (AIC) and Bayesian (BIC) information criteria as well as on Vuong tests, I found two model specifications to be well suited.

The first is a *Hurdle-at-Zero Negative Binomial Model*, also simply referred to as a two-part model, which models two distinct statistical processes for the incidence and the duration of long-term absenteeism. The first part represents the probability of crossing the hurdle, e.g., of being absent long-term, and can be estimated by a logit or probit model equivalent to that in equation (3.11). The second part models the duration of long-term absenteeism by fitting a truncated at zero Negative Binomial-2

[9] Puhani (2008) has shown that the advice of Ai and Norton (2004) to compute the discrete double difference $\frac{\Delta^2\Phi(.)}{\Delta \text{p97}\Delta \text{D}}$ is not of relevance in nonlinear models when the interest lies in the estimation of a treatment effect in a difference-in-differences model. Using treatment dummy indicators, the average treatment effect on the treated is given by $\frac{\Delta\Phi(.)}{\Delta(\text{p97*D})} = \Phi(\alpha + \beta p97_t + \gamma D_{it} + \delta DiD_{it} + s'_{it}\psi + \rho_t + \phi_s) - \Phi(\alpha + \beta p97_t + \gamma D_{it} + s'_{it}\psi + \rho_t + \phi_s\zeta)$ which is exactly what I calculate and present throughout the chapter.

(NegBin-2) model (Deb and Trivedi, 1997).

The second count data model to be employed is the so-called *Zero-Inflated Negative Binominal Model* that equally allows diverging statistical processes for the incidence and duration of long-term absenteeism. The underlying statistical mechanism differentiates between employees on long-term sick leave and those not on long-term sick leave, and assigns different probabilities that are parameterized as functions of the covariates to each group. The binary process is again specified in form of a logit or a probit model, and the count process is now modeled as an untruncated NegBin-2 model for the binary process to take on value one. Hence, zero counts may be generated in two ways: as realizations of the binary process and as realizations of the count process when the binary process is one (Winkelmann, 2008).

Both count data models incorporate the negative binomial distribution. In contrast to the more restrictive Poisson distribution, it does not only take excess zeros into account but also allows for overdispersion and unobserved heterogeneity.[10] The Negative Binomial (NegBin) Model model is a special case of a continuous mixture model. In the notation of Cameron and Trivedi (2005), the negative binomial distribution can be described as a density mixture of the following form:

$$\begin{aligned}
\varphi(y|\mu,\alpha) &= \int f\,(y|\mu,\nu) \times \gamma(\nu|\alpha)\, d\nu \\
&= \int_0^\infty \left(\frac{e^{-exp(\boldsymbol{X\beta})\nu}\{exp(\boldsymbol{X\beta})\nu\}^y}{y!}\right)\left(\frac{\nu^{\delta-1}e^{-\nu\delta}\delta^\delta}{\Gamma(\delta)}\right) d\nu \\
&= \frac{\Gamma(\alpha^{-1}+y)}{\Gamma(\alpha^{-1})\Gamma(y+1)}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\mu+\alpha^{-1}}\right)^y \qquad (3.12)
\end{aligned}$$

where $f\,(y|\mu,\nu)$ is the conditional Poisson distribution and $\gamma(\nu|\alpha)$ is assumed to be gamma-distributed with ν as an unobserved parameter with variance $\alpha = 1/\delta$. $\Gamma(.)$ denotes the gamma integral and $\mu = exp(\boldsymbol{X\beta})$ where the matrix $\boldsymbol{X}$ incorporates the same variables as the probit model in equation (3.11). The Negative Binomial Model can be derived in different ways; it has different variants and different interpretations. Note that in the special case of $\alpha = 0$ the NegBin collapses to a simple Poisson model.

[10] The unobserved heterogeneity allowed for in the NegBin-2 is based on functional form and does not capture unobserved heterogeneity which is correlated with explanatory variables.

3.5.3 Identification

In every difference-in-differences (DiD) model, the main identification assumption is the common time trend assumption. It assumes that, for both groups – treatment and control group – the trend of the outcome variable would have developed parallely in the absence of the policy intervention. In other words, after having conditioned on all available covariates, unobservables should not have a differential impact on treatment and control group with respect to changes in the dependent variable over time. Depending on the context, this may be a more or less strong assumption. My identification strategy is based on various pillars, making me confident that I am able to identify true causal reform effects.

First, I use three different subsamples that were differently affected by the two reforms. In my empirical specifications, I employ three distinct models, all of which compare these mutually exclusive subsamples to one another. The first two models contrast *Treatment Group 1* as well as *Treatment Group 2* separately to the *Control Group*, and the third model compares *Treatment Group 1* to *Treatment Group 2* (see Section 3.4.2 and Table 3.2). By this means, I esimate the net, direct, and indirect effect of the two reforms on long-term absenteeism. Comparing the findings from these three distinct models allows me to cross check the plausibility and coherence of my results.

Second, I not only estimate the reform effects on the incidence of long-term absenteeism but also the effects on the length of long-term absence spells. Working with survey data makes it possible to take a rich set of background variables into account – at the cost of having no detailed spell data. In Section 3.4.1, I have discussed why, nevertheless, the available work absence information is sufficient to measure the direct reform effect on the duration of long-term absenteeism. Moreover, I exploit an additional source of exogenous variation which allows me to distinguish effects by treatment intensity (see Section 3.2.3 for more details): The main replacement level of statutory long-term sick pay was cut along with a decrease in the upper limit of this benefit. Depending on the ratio of net to gross wages, treated employees experienced cuts of between one and ten percent of their gross wage. By using SOEP income information, I am able to calculate the individual reform induced decrease in statutory long-term sick pay remarkably exactly. I use this information in extended analyses that differentiate by treatment intensity.

Third, the implementation of the reform and the variation in the treatment intensity were clearly exogenous to the individuals and politically determined. I have not found evidence that the policy change was endogenous in the sense that the reform was a reaction to increasing absence rates (Besley and Case, 2000; German Federal Statistical Office, 2010). Rather, it was a fairly random means of cutting health expenditures and was used mainly as an instrument of the unpopular Kohl administration to demonstrate strength and capacity to act.

Fourth, as in almost every study that builds upon natural experiments, the three distinct groups that I use as control and treatment groups differ significantly in terms of their observed characteristics (see Table 3.3). For example, in comparison to the *Control Group*, *Treatment Group 1* includes fewer females but more immigrants, and the employees are less educated. *Treatment Group 2* is younger than the other subsamples, less often married, and includes more white-collar workers without tenure. The heterogeneity in most of the observable characteristics is due to the regulation of the German health insurance. However, the differences in characteristics are not the result of treatment-related self-selection but politically determined. Moreover, I adjust the sample composition with respect to all of these observed characteristics. Most importantly, I use various measures of the respondents' health status which is clearly the key determinant of long-term absenteeism. Please note that it poses no problem if the subsamples have different probabilities of being affected by long-term sickness; the identifying assumption would only be violated if unobservables existed that would impact the *change* of these probabilities differently. In case of long-term absenteeism it is unlikely that unobservables have a diverging effect on the dynamic of the outcome – after having controlled for a rich set of health-related, personal, educational, and job-related covariates as well as the annual regional unemployment rate, regional time-invariant effects, and annual time trends.

We can see from Table 3.4 that relatively few covariates affect long-term absenteeism significantly. More educated employees are less often absent for long-term periods, and firm size is positively correlated with long absence spells. As expected, the most important driver of long-term absenteeism is health status. The main reasons for long-term absences are persistently low health stocks and health shocks like unexpected illnesses and accidents (Müller et al., 1998).

Table 3.3: Variable Means by Treatment and Control Groups

Variable	Control Group	Treatment Group 1	Treatment Group 2	Min	Max
Incidence of long-term absenteeism (*longabs*)	0.033	0.060	0.026	0	1
Duration of long-term absenteeism (*longabsdays*)	1.965	3.392	2.249	0	365
Personal characteristics					
Female	0.410	0.366	0.587	0	1
Age	40.57	39.86	37.48	18	65
Age square/100	17.58	17.01	15.60	3.24	42.25
Immigrant	0.097	0.215	0.112	0	1
East Germany	0.166	0.258	0.378	0	1
Partner	0.762	0.803	0.650	0	1
Married	0.673	0.696	0.569	0	1
Children	0.483	0.470	0.435	0	1
Disabled	0.033	0.052	0.053	0	1
Good health	0.648	0.607	0.604	0	1
Bad health	0.080	0.099	0.104	0	1
No sports	0.287	0.409	0.331	0	1
Educational characteristics					
Dropout	0.021	0.050	0.044	0	1
Certificate after 8 years of schooling	0.230	0.357	0.271	0	1
Certificate after 10 years of schooling	0.290	0.330	0.438	0	1
Certificate after 12 years of schooling	0.051	0.035	0.035	0	1
Certificate after 13 years of schooling	0.363	0.115	0.162	0	1
Other degree	0.046	0.112	0.051	0	1
Work in job trained for	0.608	0.545	0.511	0	1
New job	0.204	0.179	0.179	0	1
No. of years in company	10.29	9.04	8.79	0	47.9
Job characteristics					
No tenure	0.106	0.051	0.273	0	1
One-man company	0.099	0.000	0.000	0	1
Small company	0.327	0.274	0.169	0	1
Medium-sized company	0.179	0.312	0.281	0	1
Large company	0.126	0.221	0.290	0	1
Very large company	0.268	0.193	0.260	0	1
Self employed	0.308	0.000	0.000	0	1
Blue collar worker	0.112	0.528	0.190	0	1
White collar worker	0.150	0.472	0.579	0	1
Public sector	0.493	0.000	0.829	0	1
Civil servant	0.395	0.000	0.031	0	1
Self employed	0.307	0.000	0.000	0	1
High job autonomy	0.506	0.160	0.152	0	1
Gross income per month	2,383	2,013	1,675	204	40,903
Regional unemployment rate	11.49	12.04	13.07	7	21.7
N	2,693	16,006	6,500		

Table 3.4: Probit Model: Determinants of the Incidence of Long-Term Absenteeism

Variable	Coefficient	Standard Error
Personal characteristics		
Female (d)	-0.001	0.003
Age	0.000	0.003
Age squared/100	0.000	0.001
Immigrant (d)	0.004	0.005
East Germany (d)	-0.012	0.011
Partner (d)	0.006	0.004
Married(d)	-0.008*	0.005
Children (d)	-0.006**	0.003
Disabled (d)	0.034***	0.007
Good health (d)	-0.026***	0.003
Bad health (d)	0.076***	0.007
No sports (d)	0.007**	0.003
Educational characteristics		
Certificate after 8 years' of schooling (d)	-0.006	0.006
Certificate after 10 years' of schooling (d)	-0.008	0.007
Certificate after 12 years' of schooling (d)	-0.018***	0.007
Certificate after 13 years' of schooling (d)	-0.013**	0.006
Other certificate (d)	-0.003	0.007
Work in job trained for (d)	-0.001	0.003
New job (d)	0.006	0.004
No. of years in company	-0.000	0.000
Job characteristics		
No tenure last year (d)	-0.009**	0.004
Medium-sized company (d)	0.0012***	0.004
Large company (d)	0.015***	0.004
Very large company (d)	0.014**	0.005
White collar worker (d)	-0.013***	0.003
High job autonomy (d)	-0.008*	0.004
Gross wage per month/1000	-0.005**	0.002
Regional unemployment rate	0.003	0.002
Year 1996 (d)	0.004	0.004
Year 1997 (d)	-0.004	0.006
Year 1998 (d)	-0.000	0.005
R-squared	0.106	
χ^2	916.944	
N	25199	

* p<0.10, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. (d) for discrete change of dummy variable from 0 to 1. Marginal effects, which are calculated at the means of the covariates, are displayed. Dependent variable: dummy that is 1 if respondent had long-term absence spell (*longabs*). Probit model is estimated. Regression includes state dummies. Left out reference categories are dropout, blue collar worker, and small company.

Fifth, to prove the consistency of the results, I perform various robustness checks. Thanks to the panel structure of my data, I am able to control for labor force and panel attrition by using balanced panels. Moreover, I experiment with different pre- and post-reform years. Additionally, to assess whether effect heterogeneity plays a role, I restrict the sample to singles, persons aged 25 to 55 employed full-time, and split the sample at the median wage.

In recent years, there has been an extensive debate about the drawbacks and limitations of DiD estimation. A particular concern is the underestimation of OLS standard errors due to serial correlation in case of long time horizons and unobserved (treatment and control) group effects. To deal with the serial correlation issue, I focus on short time horizons. As Bertrand et al. (2004) have shown, one main source for understating the standard errors stems from serial correlation of the outcome and the intervention variable and is basically eliminated when focusing on less than five periods. While there is consensus about the serial correlation problem, the issue with unobserved common group effects is still a matter of considerable debate. If one takes the objection of Donald and Lang (2007) seriously, then it would not be possible to draw inferences from DiD analyses in the case of few groups, meaning that no empirical assessment could be performed. I subscribe to the view of Wooldridge (2006), who says of the study by Donald and Lang (p. 18): *"DL criticize Card and Krueger (1994) for comparing mean wage changes of fast-food workers across two states because Card and Krueger fail to account for the state effect (New Jersery or Pennsylvania) [...]. But the DL criticism in the $G = 2$ case is no different from a common question raised for any difference-in-differences analyses: How can we be sure that any observed difference in means is due entirely to the policy change? To characterize the problem as failing to account for an unobserved group effect is not necessarily helpful."*[11] Besides focusing on short time spans to resolve serial correlation concerns, I use robust standard errors and correct for clustering at the individual level throughout the analysis.

Finally, as an important feature of this chapter, I can exclude that selection into or out of the treatment drives the results, which is a central issue in other settings, e.g., when labor market programs are evaluated. The reason lies in the institutional setting: Switching between the two diverse health care systems – remember that only employees insured with the SHI were affected by the cut in statutory long-term sick

[11] In this very readable extended version of an older published AER paper (Wooldridge, 2003), Wooldridge (2006) discusses several other shortcomings and assumptions of the estimation approach proposed by Donald and Lang (2007).

pay – is not allowed for the great majority. I am able to identify the only subsample that has this right to opt out of the SHI and exclude it in my robustness checks.[12]

My basic empirical strategy is to pool the data for the years 1995 to 1998 and to estimate various diffference-in-differences models. As explained above, using different subsamples which I compare against each other, I run three main models to estimate the net, the direct, and the indirect effect of the sick pay reforms on long-term absenteeism. In addition, in extended models, I differentiate by treatment intensity. Moreover, I do not only estimate the effects on the incidence of long-term absenteeism but also on the duration of long-term absenteeism.

3.6 Results

3.6.1 Assessing the Causal Reform Effect on Long-Term Absenteeism

Table 3.5 provides the unconditional DiD estimates of the reforms' net and direct effects on the incidence of long-term absenteeism. The unconditional long-term absence incidence for *Treatment Group 1* decreased from 6.16 percent in the pre-reform years 1995/1996 to 5.92 percent in the post-reform years 1997/1998. The incidence for *Treatment Group 2* decreased from 3.77 to 3.56 percent. Without the availability of a control group and by means of before-after estimators one could erroneously attribute the total decrease to the reform. However, the incidence for the *Control Group* also decreased from 3.49 to 3.11 percent in the same time period, resulting in overall difference-in-differences (DiD) estimates of +0.13 and +0.17 percent, respectively.

Table 3.6 shows the same estimates for the duration of long-term absence spells. The average number of long-term sick leave benefit days decreased between the pre- and the post-reform period from 3.62 to 3.17 days for *Treatment Group 1* and from 2.58 to 1.95 days for *Treatment Group 2*. It also decreased slightly from 1.98 to 1.95

[12] Only employees who are optionally insured with the SHI (self-employed, civil servants, and high-income earners above the income threshold) have the right to opt out of the SHI and to become part of the PHI (see Section 3.2). However, it is very unlikely that employees opted out of the SHI as a reaction to the cut in statutory long-term sick pay. Opting out is a lifetime decision since switching back to the SHI system is almost impossible. Moreover, the elderly would have to pay extremely high premiums and it makes no sense for the young either, since they are very likely to be unaffected by long-term absenteeism anyway.

days for the *Control Group* leading to unconditional DiD estimates of -0.42 and -0.61 days.

Table 3.5: Unconditional DiD Estimates on the Incidence of Long-Term Absenteeism

	1995/1996	**1997/1998**	**Difference**	**Diff-in-Diff**
Treatment Group 1	0.0616	0.0592	-0.0024	0.0013
	(0.0027)	(0.0026)	(0.0038)	(0.0078)
Treatment Group 2	0.0377	0.0356	-0.0020	0.0017
	(0.0034)	(0.0032)	(0.0047)	(0.0082)
Control Group	0.0349	0.0311	-0.0038	
	(0.0049)	(0.0048)	(0.0069)	

Average incidence rate of long-term absenteeism (*longabs*) is displayed. Standard errors in parentheses.

Table 3.6: Unconditional DiD Estimates on the Average Number of Long-Term Sick Leave Benefit Days

	1995/1996	**1997/1998**	**Difference**	**Diff-in-Diff**
Treatment Group 1	3.6212	3.1747	-0.4464	-0.4219
	(0.2455)	(0.2277)	(0.3344)	(0.7358)
Treatment Group 2	2.5800	1.9461	-0.6339	-0.6094
	(0.3407)	(0.2689)	(0.4304)	(0.7836)
Control Group	1.9767	1.9522	-0.0245	
	(0.4194)	(0.4546)	(0.6177)	

Average number of long-term absent benefit days (*longabsdays*) is displayed. Standard errors in parentheses.

Table 3.7: Difference-in-Differences Estimates on the Incidence of Long-Term Absenteeism

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
DiD1	**0.0035**	**0.0024**	**0.0053**	**0.0032**	**0.0061**	**0.0063**
	(0.0119)	**(0.0108)**	**(0.0101)**	**(0.0104)**	**(0.0088)**	**(0.0086)**
Post-reform dummy	-0.0012	-0.0123	-0.0133	-0.0102	-0.0117	-0.0102
(p1997)	(0.0124)	(0.0140)	(0.0140)	(0.0135)	(0.0127)	(0.0123)
Year 1996	0.0064	-0.0002	0.0003	0.0003	-0.0007	0.0001
	(0.0048)	(0.0053)	(0.0052)	(0.0052)	(0.0047)	(0.0047)
Year 1997	-0.0032	-0.0051	-0.0042	-0.0057	-0.0049	-0.0047
	(0.0050)	(0.0046)	(0.0045)	(0.0045)	(0.0042)	(0.0041)
Treatment Group 1	0.0276***	0.0244***	0.0151**	0.0219***	0.0145***	0.0124**
	(0.0062)	(0.0057)	(0.0063)	(0.0059)	(0.0053)	(0.0059)
Educational characteristics	no	no	yes	no	no	yes
Job characteristics	no	no	no	yes	no	yes
Personal characteristics	no	no	no	no	yes	yes
Regional unemployment rate	no	yes	yes	yes	yes	yes
State dummies	no	yes	yes	yes	yes	yes
R-squared	0.0049	0.0091	0.0308	0.0258	0.1046	0.1153
χ^2	30.368	51.609	187.191	153.235	704.315	780.916
N	18699	18699	18699	18699	18699	18699

* $p<0.1$, ** $p<0.05$, *** $p<0.01$; standard errors in parentheses are adjusted for clustering on person identifiers. Marginal effects are displayed and calculated at the means of the covariates except for *Treatment Group 1* (=1), *p1997* (=1), *Year 1996* (=0), and *Year 1997* (=1). Dependent variable: dummy that is 1 if respondent had long-term absence spell (*longabs*). Every column represents one probit model as in equation 3.11. *DiD1* is the DiD indicator. It has a one for respondents in *Treatment Group 1* in post-reform years. *DiD1* estimates the net reform effect.

The DiD estimator is now incorporated into a regression framework. Table 3.7 reports the results from six model specifications that differ with respect to the inclusion of sets of covariates and measure the reforms' net effect on the incidence of long-term absenteeism. Each specification represents a probit model equivalent to equation (3.11). The dependent variable *longabs* is 1 if the respondent had a long-term sickness spell and zero otherwise. The variable of interest is displayed as *DiD1* and is one for employees in *Treatment Group 1* in the post-reform period. In every specification, marginal effects are calculated and displayed. In none of the model specifications is the *DiD1* estimate statistically different from zero. The estimated coefficients are very close to zero, 0.0063 in the preferred specification, and positive. The standard error in the preferred specification is 0.0086. Note that the *DiD1* coefficients are robust to the inclusion of sets of covariates and close to the unconditional DiD estimate, which reinforces the plausibility of the common time trend assumption.

In the next step, I disentangle the net effect of the reform into a direct effect and an indirect effect, and estimate their impact on the incidence of long-term absenteeism separately. A priori, one would expect the sign of the direct effect to be negative since it assesses the impact of the cut in statutory long-term sick pay on long-term absenteeism. The indirect effect stems from the fact that the gap in the replacement levels between statutory short- and long-term sick leave shrank due to the reform, which might have had a positive impact on long-term absenteeism. As has been shown theoretically in Section 3.3, being able to disentangle these potentially diverging effects is important since it may be that the indirect reform effect compensated the direct effect.

Column (1) in Table 3.8 once again displays the net effect; the regression model equals Model 6 in Table 3.7. Column (2) estimates the effect of the cut in statutory long-term sick pay on the incidence of long-term absenteeism, i.e., the direct effect. In contrast to column (1), *Treatment Group 2* – those *only* affected by the cut in statutory long-term sick pay – is contrasted with the *Control Group*. The regressor of interest is now *DiD2*. The *DiD2* estimate is again positive and statistically not different from – but close to – zero. The findings from column (1) and (2) are confirmed in column (3). Here, I compare those who were affected by both reforms to those who were only affected by the cut in statutory long-term sick pay, i.e., *Treatment Group 1* to *Treatment Group 2*. Again, point estimate and standard error are close to zero in magnitude and the indirect reform effect on the incidence of long-term absenteeism is not statistically different from zero.

Table 3.8: DiD Estimation on Incidence: Disentangling the Direct from the Indirect Reform Effect

Variable	**Net effect**	**Direct effect**	**Indirect effect**
DiD1	**0.006**		
	(0.009)		
T1	0.012**		
	(0.006)		
DiD2		**0.010**	
		(0.010)	
T2		-0.015	
		(0.012)	
DiD3			**-0.000**
			(0.004)
T3			-0.021***
			(0.006)
Post-reform dummy	-0.010	0.007	-0.000
(p1997)	(0.012)	(0.012)	(0.004)
Year 1996	0.000	0.016*	0.002
	(0.005)	(0.009)	(0.003)
Year 1997	-0.005	0.009	-0.003
	(0.004)	(0.007)	(0.003)
Educational characteristics	yes	yes	yes
Job characteristics	yes	yes	yes
Personal characteristics	yes	yes	yes
Regional unemployment rate	yes	yes	yes
State dummies	yes	yes	yes
R-squared	0.115	0.106	0.114
χ^2	780.916	298.763	1074.389
N	18699	9193	22506

* $p<0.1$, ** $p<0.05$, *** $p<0.01$; standard errors in parentheses are adjusted for clustering on person identifiers. Marginal effects are displayed and calculated at the means of the covariates except for *T1 (2, 3)* (=1), *p1997* (=1), *Year 1996* (=0), and *Year 1997* (=1). Dependent variable: dummy that is 1 if respondent had long-term absence spell (*longabs*). Every column represents one probit model as in equation 3.11. *T1* is one for respondents who were affected by both cuts, in statutory short- *and* long-term sick pay (*Treatment Group 1*), and is zero for those who were not affected at all by the reforms (*Control Group*). *T2* contrasts those only affected by the cut in statutory long-term sick pay (*Treatment Group 2*) to the *Control Group*, and *T2* compares the incidence of *Treatment Group 1* with the incidence of *Treatment Group 2*. *DiD1 (DiD2, DiD3)* is the DiD indicator and has a one for *T1 (T2, T3)* =1 and post-reform years. *DiD1 (DiD2, DiD3)* estimates the net (direct, indirect) reform effect. More information about the treatment indicators can be found in Section 3.4.2 and in Appendix B.

Table 3.9: DiD Estimation on Incidence with Varying Treatment Intensity

Variable	Net effect	Direct effect
DiD1index	**0.000**	
	(0.001)	
T1index	0.003***	
	(0.001)	
DiD2index		**0.000**
		(0.001)
T2index		0.000
		(0.002)
Post-reform dummy	-0.005	0.011
(p1997)	(0.010)	(0.012)
Year 1996	0.000	0.016*
	(0.005)	(0.009)
Year 1997	-0.005	0.009
	(0.004)	(0.007)
Educational characteristics	yes	yes
Job characteristics	yes	yes
Personal characteristics	yes	yes
Regional unemployment rate	yes	yes
State dummies	yes	yes
R-squared	0.116	0.104
χ^2	785.887	291.684
N	18699	9193

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. Marginal effects are displayed and calculated at the means of the covariates except for *p1997* (=1), *Year 1996* (=0), and *Year 1997* (=1). Every column represents one probit model as in equation 3.11. Dependent variable: dummy that is 1 if respondent had long-term absence spell (*longabs*). *T1index (T2index)* is the treatment intensity index for the same subsamples as *T1 (T2)*. It takes on positive values on a continuous scale up to 10.00 for *Treatment Group 1* and is zero for the *Control Group*. *DiD1index (DiD2index)* is the DiD intensity index and has positive values for *Treatment Group 1 (2)* and post-reform years. *DiD1index (DiD2index)* estimates the net (direct) reform effect. More information about the treatment intensity indices can be found in Section 3.4.2 and in Appendix B.

Table 3.10: DiD Estimation on the Duration of Long-Term Absenteeism

	Zero-Inflated Model	***Hurdle-at-Zero Model***
Variable	**Direct effect: Varying Intensity**	**Direct effect: Varying Intensity**
DiD2index	**-0.041**	**-0.904**
	(0.058)	**(1.915)**
T2index	0.043	1.188
	(0.044)	(1.006)
Post-reform dummy	-0.402	-16.524
(p1997)	(0.642)	(24.307)
Year 1996	-0.064	1.509
	(0.275)	(10.047)
Year 1997	0.242	0.071
	(0.326)	(14.345)
Educational characteristics	yes	yes
Job characteristics	yes	yes
Personal characteristics	yes	yes
Regional unemployment rate	yes	yes
State dummies	yes	yes
χ^2	149.552	108.45
N	9193	327

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. Marginal effects are displayed and calculated at the means of the covariates except for *p97* (=1), *Year 1996* (=0), and *Year 1997* (=1). Every column represents one count data model as in equation 3.12. Dependent variable: Number of long-term sick leave benefit days (*longabsdays*). *T2index* is the treatment intensity index for the same subsample as *T2*. It takes on positive values on a continuous scale up to 10.00 for *Treatment Group 2* and is zero for the *Control Group*. *DiD2index* is the DiD intensity index and has positive values for *Treatment Group 2* and post-reform years. *DiD2index* estimates the direct reform effect. More information about the treatment intensity indices can be found in Section 3.4.2 and in Appendix B.

T1index and *T2index* represent the treatment intensity of the reform, which I define as the cut in statutory long-term sick pay relative to the individual's gross wage (see Section 3.2.3 and 3.4.2). By interacting these continuous variables with the post-refom dummy *p1997*, I estimate the net effect and the direct effect on the incidence of long-term absenteeism in Table 3.9. As above, I am unable to reject the hypothesis that the reforms have induced any significant behavioral changes, which is illustrated by the *DiD1index* and *DiD2index* coefficients that are very close to zero in size and not significantly different from zero.

Table 3.10 uses the number of days that long-term sick leave benefits were received (*longabsdays*) as dependent variable and estimates count data models as explained in Section 3.5.2. I always focus on the direct effect and differentiate by treatment intensity, i.e., I use *T2index* and its interaction with *p1997* (*DiD2index*). The non-significant point estimate for the whole sample is -0.041, and conditional on those who had a long-term absence spell, it is -0.904 (days).

3.6.2 Robustness Checks and Heterogeneity in Effects

Until now my estimation strategy was to pool the data over four years, which means that I allowed the sample composition to change over the years. As people with long-term absence spells have a higher probability to leave the labor force as a result of their (probably severe) illness, I should check whether this selection out of the labor market drives my results. From those who had a long-term absence spell in 1996, 7.1 percent did not answer the questionnaire one year later for unknown reasons (one respondent died and one moved abroad). I do not find evidence that long-term illness led to a higher probability of dropping out of the sample in the subsequent year, since 7.7 percent of the respondents without long-term absence spells did not participate in the following year. One the other hand, 74.6 percent of those who were absent for a long-term period in 1996 were employed full-time at that time, whereas one year later, this number decreased to 62.3 percent for those who remained in the sample. Especially if I had found reform effects that suggested a significant reduction in long-term absenteeism, the estimate might have been driven by selection out of the labor market. In the following, I discuss why illness-related selection out of the labor market is no source of serious concern in this setting.

First, in 1998 (with information about 1997) the SOEP group drew a random refreshment sample that covered all existing subsamples and a total of 1,067 observations (Wagner et al., 2007). Thanks to this refreshment sample, the employment status distribution over those who had long-term sickness spells in 1996 and 1997 remained very stable. Under the consideration of the refreshment sample, in total, 73.1 percent of those who suffered long-term absence spells in 1997 were employed full-time (as compared to 62.3 percent without considering the refreshment sample).[13]

Second, the availability of a control group allows me to control for treatment-independent selection out of the labor market.[14] In the absence of a control group one could easily confuse the illness-related selection out of the labor market with a causal reform effect, since it is natural that sickness absence rates decrease over time as the sample ages.

Finally, since I use panel data, in addition to correcting the sample composition by observables, I use a balanced sample in one of the robustness checks shown in Table 3.11 and 3.12.

Table 3.11 and 3.12 report results for the *direct* effect specification on the incidence and duration of long-term absenteeism. Both tables use *T2index* and *DiD2index*, meaning that I always differentiate by treatment intensity. As a first test, I center the data two years around the reform (column (1)). Afterwards, I restrict my sample to the years 1996 and 1997, balance it, and consider only employees who were eligible for long-term sick pay in both years and who answered the SOEP questionnaire in both years (column (2)). An alternative robustness check would be to take 1995 as reference year and contrast it with 1997 and 1998. It might have been the case that anticipation effects played a role and that employees already adapted their behavior in 1996, when the reform plans were made public (column (3)). This is, however, not very probable as many catalysts of long-term absences, like cancer diagnosis, happen unexpectedly. Since people who started their long-term absence spell in 1996 and carried it over to 1997 took advantage of a transitory arrangement and were not exposed to reduced sick pay, I contrast the years 1995/1996 with 1998 in column (4).

[13] For the other employment groups like the part-time employed, the deviation between 1996 and 1997 was less than 1.6 percent.

[14] I cannot, however, entirely exclude the possibility that the reform had an effect on the decision to leave the labor market voluntarily. I am unable to observe how large the share of voluntary labor market quitters was. However, as the cut in long-term sick pay was moderate and financial penalties are substantially higher for unemployed or retirees, reform-induced selection out of the labor market is likely to play a negligible role.

Table 3.11: Robustness and Heterogeneity of Effects: Direct Effect on Incidence Using Treatment Index 2

Variable	**'96-'97**	**'96-'97; balanced**	**'95 vs. '97/'98**	**'95/'96 vs. '98**	**Full-time: 25 - 55**	**Singles**	**No option. insured**	**<median income**	**>median income**
DiD2index	0.000	0.002	0.001	0.002	0.001	0.002	0.001	0.001	0.002
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Educational characteristics	yes	yes	yes	yes	yes	yes	yes	yes	yes
Job characteristics	yes	yes	yes	yes	yes	yes	yes	yes	yes
Personal characteristics	yes	yes	yes	yes	yes	yes	yes	yes	yes
Regional unemployment rate	yes	yes	yes	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.096	0.123	0.084	0.089	0.095	0.110	0.079	0.118	0.101
χ^2	145.022	126.841	167.372	217.029	144.648	113.32	207.033	212.115	166.736
N	4595	3239	6786	6827	5204	2747	8435	4833	4289

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. Marginal effects are calculated at the means of the covariates except for *p1997* (=1), *Year 1996* (=0), and *Year 1997* (=1). Every column represents one probit model as in equation 3.11. Dependent variable: dummy that is 1 if respondent had long-term absence spell (*longabs*). *DiD2index* is the DiD intensity index and has positive values for *Treatment Group 2* and post-reform years. It is zero for respondents in the *Control Group*. *DiD2index* estimates the direct reform effect. More information about the treatment intensity indices can be found in Section 3.4.2 and in Appendix B.

Table 3.12: Robustness and Heterogeneity of Effects: Direct Effect on Duration Using Treatment Index 2

Variable	**'96-'97**	**'96-'97; balanced**	**'95 vs. '97/'98**	**'95/'96 vs. '98**	**Full-time: 25 - 55**	**Singles**	**No option. insured**	**<median income**	**>median income**
DiD2index	-0.021	0.130	-0.035	-0.025	-0.041***	0.063	-0.093	-0.114**	-0.048
	(0.053)	(0.123)	(0.039)	(0.024)	(0.020)	(0.072)	(0.071)	(0.023)	(0.049)
Educational characteristics	yes	yes	yes	yes	yes	yes	yes	yes	yes
Job characteristics	yes	yes	yes	yes	yes	yes	yes	yes	yes
Personal characteristics	yes	yes	yes	yes	yes	yes	yes	yes	yes
Regional unemployment rate	yes	yes	yes	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	
χ^2	4608.620	1933.945	5256.873	2111.791	2478.681	222.277	235.314	2332.530	6751.009
N	4571	3334	6786	6812	5186	2798	8435	4833	4289

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. Marginal effects are calculated at the means of the covariates except for *p1997* (=1), *Year 1996* (=0), and *Year 1997* (=1). Every column represents one Zero-Inflated NegBin-2 model as in equation 3.12. Dependent variable: number of long-term sick leave benefit days (*longabsdays*). *DiD2index* is the DiD intensity index and has positive values for *Treatment Group 2* and post-reform years. It is zero for respondents in the *Control Group*. *DiD2index* estimates the direct reform effect. More information about the treatment intensity indices can be found in Section 3.4.2 and in Appendix B.

To test effect heterogeneity, I restrict the sample to full-time employed people aged 25 to 55 (column (5)) and to singles (column (6)) as the income of other household members may have had an impact on the exposure to treatment. On the household level, the relevant parameter might be the decrease in total household income rather than in individual wages. Since optionally SHI insured could have switched to the PHI system as a reaction to the reform, I exclude all optionally insured people in column (7). I also split the sample at the median gross wage (columns (8) and (9)).

Table 3.11 shows the results when I use the incidence of long-term sick leave, *longabs*, as dependent variable. None of the *DiD2index* coefficients is statistically different from zero but all are very close to zero in magnitude, which reinforces my main findings above. Note that although all coefficients are practically zero, they all have positive signs.

In Table 3.12, where I use the number of long-term sick leave benefit days (*longdaysabs*) as dependent variable, I do not find significant reform effects for most of the specifications either. The coefficients are close to zero in magnitude, and columns (2) and (6) even have positive signs. However, I find significantly negative reform effects for middle-aged full-time employed and the poor (columns (5) and (8)), which suggests heterogeneity in the reform effects on benefit duration. According to the estimates, a one unit increase in *T1index*, which equals an increase in the absence costs of about 5 percent, led to a decrease in the average number of long-term sick leave benefit days of around 0.04 and 0.11, respectively. Middle-aged full-time employed people most likely need to feed a family and might be the main earners in their household. The poor are also likely to be more crucially dependent on their full salary. Besides the notion that these subsamples have reacted to monetary incentives, another explanation is possible: Although *Treatment Group 2*, which I use in these specifications, was solely affected by the cut in statutory long-term sick pay, there might have been spillover effects from the cut in statutory short-term sick pay. Since Puhani and Sonderhof (2010) has shown that the cut in statutory short-term sick pay clearly induced reductions in short-term sick pay, it is at least imaginable that public sector employees and trainees insured with the SHI were not fully aware of their priviledges. If the cut in statutory short-term sick pay reduced short-term sickness spells that these groups might have had *in addition* to their long-term spell, my estimates here would be contaminated. Moreover, it might have been the case that spillover effects were induced if employees in *Treatment Group 1* had partners working in the private sector who reacted to the cut in short-term sick pay.

A conventional method for checking the robustness of DiD estimates is to perform placebo regressions and to estimate the reform effects for years without a reform. For the assumption of common time trends of control and treatment group to hold, none of the placebo reform effects should be significant. Table 3.13 displays placebo regression results on the incidence and duration of long-term absenteeism for the years 1995 and 1996. All placebo estimates turn out to be insignificant.

Table 3.13: Placebo Estimates Using Treatment Index 2

Variable	**Direct effect (Incidence)**	**Direct effect (Duration)**
DiD2index96	0.001	-0.042
	(0.003)	(0.159)
DiD2index95	-0.003	-0.171
	(0.005)	(0.277)
Educational characteristics	yes	yes
Job characteristics	yes	yes
Personal characteristics	yes	yes
Regional unemployment rate	yes	yes
State dummies	yes	yes
χ^2	339.092	264.462
N	11457	11457

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are adjusted for clustering on person identifiers. Marginal effects are calculated at the means of the covariates except for corresponding post reform dummies (=1), pre-treatment(=0), and post-treatment years (=1). Column (1) estimates one probit model as in equation 3.11 and column (2) estimates one Zero-Inflated NegBin-2 model as in equation 3.12. Dependent variable in column (1): incidence of long-term absenteeism (*longabs*). Dependent variable in column (2): number of long-term benefit days (*longabsdays*). *DiD2index96 (95)* is the DiD intensity index for a pseudo-reform in 1996 (1995) and has positive values for pseudo-*Treatment Group 2* and pseudo-post-reform years. It is zero for respondents in the pseudo-*Control Group*. *DiD2index96 (95)* estimates the pseudo direct reform effect. More information about the treatment intensity indices can be found in Section 3.4.2 and in Appendix B.

3.6.3 Calculation of SHI Reform Savings

Statutory long-term sick pay amounted to 80 percent of the monthly gross wage before the reform and was reduced to 70 percent after the reform. The benefit cap decreased from 100 percent of the monthly net wage to 90 percent of the wage after taxes and social contributions. I calculate the total price adjusted SHI reform savings from 1997 to 2006, reflecting the redistributional effect of the reform. Reducing the sick pay level for the long-term sick benefited the rest of the statutory health insurance pool through lower contribution rates.

As a first estimate, I calculate statutory long-term sick pay according to the old and the new regulations for every eligible individual and the years 1997 to 2006, take the difference, and sum over the frequency-weighted number of long-term absences for the whole period.

Through the reform, statutory long-term sick pay has been cut on average by approximately €300 per case and year. Since (reduced) social contributions are charged on long-term sick pay, the net cut per case was about €250. Given that the average number of long-term sick leave benefit days equals about 2.5 months, this translates into a benefit cut of about €100 per month. The decrease represents about seven percent of the average monthly net wage.

Comparing the frequency-weighted number of SHI long-term sickness cases in the SOEP with the administrative data reveals that the SOEP underestimates the number of cases as well as the average benefit days per case. This is not surprising since long-term sick people with very long sickness spells have a particularly high probability of not participating in the survey.

Consequently, I make use of administrative data from the German Ministry of Health on the total number of SHI long-term sick pay cases and the average number of long-term sick leave benefit days for SHI insured. Unfortunately, no personal data and no income information are collected by official statistics. Hence, I combine administrative data with the SOEP data set, which contains very detailed income information. By this means, I estimate that, between 1997 and 2006, the total sum that the SHI saved due to the reform amounted to around €5.5 billion.[15] Considering social contributions, this translates into an accumulated net loss for the long-term sick of about five billion euros during that period of time.

[15] In the working paper version, I present a more detailed analysis of the redistributional effects. Various specifications and different scenarios are discussed.

3.7 Discussion and Conclusion

To the best of my knowledge, this is the first study that explicitly analyzes how cuts in statutory long-term sick leave affect long-term absenteeism in the context of the European statutory sickness absence insurance. In the first part of the chapter, by means of a simple dynamic model of absence behavior, I analyze the different incentive effects on long-term absenteeism that were triggered by two cuts in the German statutory sick pay scheme. However, under the assumption that employees on long-term sick leave are seriously sick, the incentive structure of the sick pay scheme breaks down and employees would not react to monetary incentives.

In the second part of the chapter, I use SOEP panel data to estimate the reform effects on long-term absenteeism empirically. This is feasible by means of conventional difference-in-differences models since the cut in stautory long-term sick pay applied universally to all employees insured with the public health insurance, but not to respondents insured with the private health insurance. In Germany, the two health care systems co-exist independently. Since switching between the two systems is almost impossible due to federal regulations, I can exclude that treatment-related selection drives my results. Moreover, the reform was clearly exogenous to the individual and a fairly random instrument of the ruling administration to cut health expenditures and to demonstrate capacity to act.

I run three distinct main difference-in-differences model that all contrast mutually exclusive subsamples that were differently affected by the reforms with one another. Moreover, I do not only estimate the effects on the incidence of long-term absenteeism but also on the number of long-term sick leave benefit days. In addition, I am able to differentiate by treatment intensity since one element of the reform induced additional exogenous variation such that employees were not affected equally by the reform.

The consistency of the findings from this variety of approaches, together with the results from various robustness checks, makes me confident of having identified true causal reform effects which are not driven by diverging time trends or selection effects. All empirical models consistently suggest that the incidence of long-term absence spells was not affected by the cut in statutory long-term sick pay. All effects on the incidence are close to zero in magnitude and even have positive signs. This suggests that not only imprecision in the estimates leads to the conclusion that employees have not adapted their long-term sick leave behavior. As for the effects on the duration of long-term absenteeism, I also find mostly insignificant reform effects but the sign of the effects is negative, as expected. However, for two subsamples – the poorer half of the sample as well as middle-aged employees working full-time – I find that the cut in statutory long-term sick pay reduced the length of these spell significantly. This suggests heterogeneity in the reform effects on the number of benefit days.

My empirical results are in line with the findings from Campolieti (2004) who convincingly showed that benefit recipients of the Candadian disability insurance (DI) have not adapted their labor supply behavior as a reaction to changes in benefits. Moreover, the results are also partly in line with the findings from Curington (1994), who used U.S. data from the 1970s on the workers' compensation insurance (WCI). However, the European statutory sickness absence insurance is not directly comparable to the North American DI and WCI.

I have several explanations for my main finding that the long-term sick have not significantly adapted their sick leave behavior to benefit changes: First, the result is in line with my model predictions if long-term sick people are assumed to be seriously ill. This is plausible since, in Germany, the most common causes for sickness spells of more than six weeks are chronic diseases of the spine, arthritis, accidents, cancer, and mental diseases. Moreover, 43 percent of the persons concerned have strong or very strong fears of being laid off and becoming unemployed (Müller et al., 1998). The causes for long-term absenteeism differ substantially from those for short-term abenteeism. Short-term sick leave is mostly determined by flus and light illnesses which clearly leave more space for moral hazard, especially when physicians' certificates are not required during the first days of a spell.

Second, the stringency of the sick leave monitoring and screening process is a potentially important determinant of labor supply reactions and moral hazard. In Germany, certification requirements increase with the length of spells. After six weeks of continuous sick leave, physicians need to issue different certificates in regular time intervals. Moreover, as discussed in Section 3.2.2, German social legislation explicitly requires the Medical Service of the Statutory Health Insurance (SHI) to take measures that prevent long-term absenteeism and the risk of patients descending the social ladder through long-term illness. Likewise, both employers and sickness funds have clear incentives to avoid unnecessary and overlong sickness episodes. They are encouraged by law to cooperate with the Medical Service which employed about 2,000 independent physicians and examined 1.7 million cases of absenteeism in 2007

(Medizinischer Dienst der Krankenversicherung (MDK), 2008).

Third, relative to short absence periods, the incentive structure of the German statutory sick leave scheme makes long absence periods unattractive. The replacement level does not increase with the duration of a spell – as in other European countries like Spain, Czech Republic, or Portugal – but decreases. As has also been shown theoretically in Section 3.3, decreasing benefits yields an incentive to accumulate shirking behavior in the lower end of the sickness spell distribution rather than in the upper end.

Finally, given that sickness episodes of more than six weeks are typically triggered by serious sickness, the cut in long-term sick pay may have been too moderate to induce changes in the labor supply behavior. My calculations suggest that, on average, the long-term sick have lost €250 per spell or €100 per month – the latter figure represents seven percent of the monthly net wage.

By combining SOEP income data with administrative data, I estimate that the cut in statutory long-term sick pay redistributed five billion Euros from the long-term sick to the SHI insurance pool in order to achieve lower contribution rates. This was the reform's main objective: cutting health expenditures in order to achieve lower contribution rates and to stimulate job creation.

Various pieces of evidence throughout this chapter suggest that moral hazard is of minor importance when sickness spells of more than six weeks are considered. Consequently, health reforms like the German one do not lead to more efficient sickness insurance markets by decreasing the degree of moral hazard but are merely an instrument to cut health expenditures. On the other hand, when cuts in replacement levels are moderate, this cost containment instrument seems to be economically efficient in the sense that it induces no major behavioral reactions that might lead to undesirable equilibria. Policy makers should be aware of the redistributional consequences. It is simply a normative question whether such an instrument to cut health expenditures should be applied.

The U.S. and Canada have no social insurance comparable to the European statutory sickness insurance. However, the number of DI recipients is growing steadily in these countries, imposing large economic and social costs since DI recipients usually completely withdraw from the labor market. Moreover, various studies suggest that moral hazard plays a crucial role in the U.S. DI insurance program. The main idea of the European statutory sickness insurance is to provide relatively generous benefits even for work-*un*related sickness while keeping employees employed, together with a stringent monitoring system. This might be a more promising approach to maintain employees' working capacity in the long run. Further research on this topic is needed.

Appendix B

Table 3.14: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Longabs	0.049	0.215	0	1	25,199
Longabsdays	2.967	19.449	0	335	25,199
T1	0.856	0.351	0	1	18,699
T2	0.707	0.455	0	1	9,193
T3	0.289	0.453	0	1	22,506
T1index	5.699	2.755	0	10	18,699
T2index	4.652	3.32	0	10	9,193
Personal characteristics					
Female	0.427	0.495	0	1	25,199
Age	39.322	11.154	18	65	25,199
Age squared/100	16.707	9.067	3.24	42.25	25,199
Immigrant	0.176	0.381	0	1	25,199
East Germany	0.28	0.449	0	1	25,199
Partner	0.759	0.428	0	1	25,199
Married	0.661	0.473	0	1	25,199
Children	0.463	0.499	0	1	25,199
Disabled	0.05	0.218	0	1	25,199
Good health	0.611	0.488	0	1	25,199
Bad health	0.098	0.298	0	1	25,199
No sports	0.376	0.484	0	1	25,199
Educational characteristics					
Drop out	0.045	0.208	0	1	25199
Certificate after 8 years' of schooling	0.321	0.467	0	1	25,199
Certificate after 10 years' of schooling	0.354	0.478	0	1	25,199
Certificate after 12 years' of schooling	0.037	0.188	0	1	25,199
Certificate after 13 years' of schooling	0.154	0.361	0	1	25,199
Other certificate	0.089	0.285	0	1	25,199
Work in job trained for	0.543	0.498	0	1	25,199
New job	0.182	0.386	0	1	25,199
No. years in company	9.106	9.217	0	47.9	25,199
Job characteristics					
No tenure	0.114	0.318	0	1	25,199
One man company	0.011	0.104	0	1	25,199
Small company	0.253	0.435	0	1	25,199
Medium-sized company	0.289	0.454	0	1	25,199
Large company	0.229	0.42	0	1	25,199

Continued on next page...

... *Table 3.14 continued*

Variable	Mean	Std. Dev.	Min.	Max.	N
Very large company	0.218	0.413	0	1	25,199
Blue collar worker	0.396	0.489	0	1	25,199
White collar worker	0.465	0.499	0	1	25,199
Public sector	0.267	0.442	0	1	25,156
Civil servant	0.05	0.218	0	1	25,199
Self-employed	0.033	0.178	0	1	25,199
High job autonomy	0.195	0.396	0	1	25,199
Gross wage per month	1965.35	1106.54	204.00	40903.35	25,199
Regional unemployment rate	12.25	3.97	7	21.7	25,199

Chapter 4

Estimating Price Elasticities of Convalescent Care Programs

Published in The Economic Journal, 120(545): 816-844

Abstract

This study is the first to estimate price elasticities of demand for convalescent care programs. In 1997, the German legislature more than doubled the daily copayments for the publicly insured from €6 to €13. The measure caused the overall demand for convalescent care treatments to fall by 20 to 25 percent. I estimate the price elasticity for medical rehabilitation programs aimed at preventing work disability to be about -0.3, whereas the elasticity for medical rehabilitation programs for recovery from accidents at work lies around -0.5. The demand for preventive treatment at health spas is elastic and less than -1.

4.1 Introduction

How does the demand for medical care change when prices change? Since the early days of the profession of health economics, the question of the price elasticity for health services has been central. Joseph P. Newhouse, one of the founders of the field of health economics, has published more than 100 articles alone on health care demand. Newhouse was also the leader of the RAND group that designed and directed the famous RAND Health Insurance Experiment that focused on the impact of cost-sharing on the demand for medical care. Not only Newhouse but also many other economists have devoted a great deal of attention to this topic.

The RAND Health Insurance Experiment (HIE) remains the largest health policy study in U.S. history to this day. It was set up in 1971 and is still the methodological "gold standard" when it comes to demand for health care. Families at six different sites in the U.S. were randomly assigned to 14 different health insurance plans and observed for up to five years (Manning et al., 1987). Various authors have highlighted different aspects of the experiment that are documented in more than 300 publications, most of them from the 1980s (Zweifel and Manning, 2000).

Empirical evidence from countries other than the U.S. is scanty at best. There have been remarkably few studies in recent years deriving price elasticities of demand for health care (Chiappori et al., 1998; Cockx and Brasseur, 2003; Bishai et al., 2008; Meyerhoefer and Zuvekas, 2010). In addition, there is surprisingly little evidence on the price elasticity of demand for preventive care, despite its potentially high relevance for public health. The few existing studies suggest that elasticities are higher for preventive care than for other medical services (Roddy et al., 1986; Keeler and Rolph, 1988; Ringel et al., 2002). The health economics literature shows a growing interest in preventive care (Kenkel, 2000; Deb, 2001; Herring, 2010).

Most studies estimate the overall price elasticity of the demand for health services to lie at around -0.2. Outpatient care is found to be more elastic than inpatient care, and mental health care more price-responsive than outpatient care. There is evidence that price elasticities are higher in the short run than in the long run. Moreover, it has been shown empirically and theoretically that elasticities are lower for more severe illnesses and for more urgent care (O'Grady et al., 1985; Wedig, 1988; Keeler et al., 1988; Lee, 1995; Zweifel and Manning, 2000).

This chapter evaluates how an increase in the copayment rate affected the demand for the two main types of convalescent care in Germany: medical rehabilitation therapy (*medizinische Rehabilitationsmaßnahmen*) and preventive therapy (*medizinische Vorsorgeleistungen*). From 1997 on, the daily copayment rates for medical rehabilitation therapy and preventive therapy more than doubled for people insured under the German Statutory Health Insurance (SHI) system. Except for the daily copayment, the SHI covers both forms of therapy in full if prescribed by a physician and carried out in an authorized medical facility in a spa town. The first type of convalescent care, medical rehabilitation therapy, consists of treatments to help patients recover from a severe illness or an accident.

With the second type of treatment, preventive therapy, the German social legislation states that to be eligible, patients need not suffer from a specific illness but merely be at risk of becoming sick in the foreseeable future without treatment. In contrast to ordinary spa vacations, patients who receive preventive therapy at SHI health spas are required to follow a strict daily schedule and take part in nutrition, calisthenics, or stress reduction programs.

Though the programs evaluated here are country-specific, similar types of treatment exist in other countries as well. For example, German preventive therapies can be seen as a form of care that educates people to improve their health behavior by developing better diet and exercise habits. I refer to both types of programs – medical rehabilitation therapy and preventive therapy – with the umbrella term *convalescent care*, which is equivalent to the German expression *Kur*. In Germany in 1995, taking all convalescent care programs together, 1.9 million patients were treated and more than €7 billion (0.4 percent of GDP) were spent on these programs (German Federal Statistical Office, 2010).

In this chapter, I contribute to the existing literature in a number of ways. First, I analyze how doubling the daily copayment rate affected the overall demand for convalescent care programs, and then I estimate how the demand for specific programs reacted. Second, this is one of the few European studies on health care demand, and it uses recent and representative data for Germany, the largest country in Europe. Third, it is noteworthy that selection into or out of the treatment is not an issue in this context. The majority of German citizens are insured compulsorily under the SHI system, which provides free universal health care coverage. Germany also has a variety of independent Private Health Insurance (PHI) providers for people in particular income and occupational groups. Strict legal regulations prevent switching back and forth between the SHI and the PHI and thus make it almost impossible to switch out of the statutory system to avoid negative consequences of health care reforms. The rigidity of the German context guarantees a control group for the evaluation of this natural experiment. Finally and perhaps most importantly, this is the first study that provides price elasticity estimates for medical rehabilitation therapies and preventive therapies.

In all cases, I analyze the effect of the increase in the copayment rate on the incidence of the program since no microdata on the length of the treatments are available. However, the average length of treatment is regulated by public law and deviations from it are determined by the medical personnel and the SHI sickness fund.

The doubling of copayments has led to a decrease in the overall incidence of convalescent care programs by 20 to 25 percent. I estimate the price elasticity for medical rehabilitation programs that aim at preventing work disability to lie at around -0.3. Moreover, I estimate the elasticity for medical rehabilitation programs due to work accidents to be about -0.5. The elasticity of demand for SHI preventative care is much more price-responsive and is likely to be less than -1. To my knowledge, this is also the first attempt to compute price elasticities for preventive care, which, according to my estimates, is price-elastic.

In the next section, I describe some features of the German health care system and give more details about the reform. In Section 4.3, the dataset and the variables used are explained, and in the subsequent section, I specify my estimation and identification strategy. Estimation results are presented in Section 4.5 and I conclude with Section 4.6.

4.2 The German Health Care System and the Policy Reform

The German health care system is actually comprised of two independent health care systems existing side by side. The more important of the two is the Statutory Health Insurance (SHI), which covers about 90 percent of the German population. Employees whose gross income from salary is below a defined income threshold (2009: €4,050 per month) are compulsorily insured under the SHI. High-income earners who exceed that threshold as well as self-employed people have the right to choose between the SHI and private health insurance. Non-working spouses and dependent children are covered at no cost by the SHI family insurance. Special regulations apply to particular groups such as students and the unemployed, but most of these are SHI-insured. Everyone insured under the SHI is subject to a universal benefit package which is determined at the federal level and codified in the Social Code Book V (SGB V). Coinsurance rates[1] are prohibited in the SHI and thus, apart from copayments, health services are fully covered. The SHI is one pillar of the German social legislation (German Ministry of Health, 2008).

The SHI is primarily financed by mandatory payroll deductions that are not risk-related. For people in gainful employment, these contributions are split equally between employer and employee up to a contribution ceiling (2009: €3,675 per month). Despite several health care reforms that have tried to remedy the problem of rising health care expenditures, contribution rates rose from 12.6 percent in 1990 to 14.9 percent in 2009 mainly due to demographic changes, medical progress, and system inefficiencies (German Federal Statistical Office, 2010).

The second track of the German health care system is Private Health Insurance (PHI). Private insurance providers primarily cover private-sector employees above the aforementioned income threshold, public-sector employees[2], and the self-employed. Privately insured people pay risk-related insurance premiums determined by an initial health checkup. The premiums exceed the expected expenditures in younger age brackets, since health insurance providers build up reserves for rising expenditures with increased age. Coverage is provided under a range of different health plans, and insurance contracts are subject to private law. Consequently, in Germany, public

[1] Coinsurance rates are important for private health insurance providers. They differ from copayments. While a copayment is typically a fixed amount that the insured person has to pay per day of treatment or for specific medical devices or medications, a coinsurance rate defines a percentage of the costs that an insured person has to pay when using the system. For example, private health insurance providers may offer 80/20 health plans specifying that the insured person has to pay 20 percent of all costs incurred while the health insurance provider is responsible for the remaining 80 percent. Often health insurance providers limit the total amount that an individual has to spend out-of-pocket with a so called coinsurance-cap which might be €2,000 per year.

[2] We need to distinguish between two types of employees in the German public sector. First, there are civil servants with tenure (called *Beamte*), henceforth called "civil servants." Most are PHI-insured since the state reimburses 50 percent of their health expenditures (*Beihilfe*) and almost all of them purchase private insurance to cover the other 50 percent of expenditures not covered by the state. Second, we need to consider employees in the public sector without legal tenure. (called *Angestellte im öffentlichen Dienst*). They have some privileges as well, but most are insured under the SHI (under the same conditions as everyone else). I refer to them here as "public servants."

health care reforms apply only to the SHI, not to the PHI.

It is important to keep in mind that compulsorily insured persons have no right to choose the health insurance system or benefit package. They are compulsorily insured under the standard SHI insurance scheme. Once an optionally insured person (a high-income earner, self-employed person, or civil servant) opts out of the SHI system, it is practically impossible to switch back into it. Employees above the income threshold are legally forbidden from switching back, while employees who fall below the income threshold in subsequent years may do so under certain conditions, but are not able to carry along the reserves that their PHI providers have built up since these are not portable (neither between PHI and SHI, nor between the different private health insurance providers).[3] In reality, switching to a private health insurance provider may be regarded as a lifetime decision, and switching between the SHI system and PHI – as well as between PHI providers – is therefore very rare. According to SOEP data, only 1.6 percent of those who were insured under the SHI for at least one year switched to the PHI between 1994 and 1998. The rate did not increase after the reform. Only 1.3 percent of those who were insured under the SHI in 1995 switched to the PHI in 1997 or 1998.

4.2.1 The German Market for Convalescent Care

In Europe, and especially in Germany, there is a long tradition of convalescent care provided in health spas to recover from poor health. Since the time of the Roman Empire, doctors have sent patients to "take the waters" to recover from various disorders. In Germany, convalescent care treatments are usually combined with various types of physical therapy, often including electrotherapy, massage, underwater exercise, ultrasonic therapy, health and diet education, stress reduction therapy, and cold and hot baths as well as mud packs. While spa therapies are usually prescribed to people suffering generally poor health and are often largely preventive, medical rehabilitation implies recovery from a specific illness or accident. Both forms of therapy require the patients to follow a strict daily schedule.

The German SHI is one of the few health insurance systems worldwide that, apart from small copayments, fully covers medical rehabilitation therapies and preventive

[3] Until 2009, accrued reserves for rising health expenditures with increased age were not portable at all. From January 1, 2009 on, portability of accrued reserves between PHI providers has been made compulsive to a strictly defined extent.

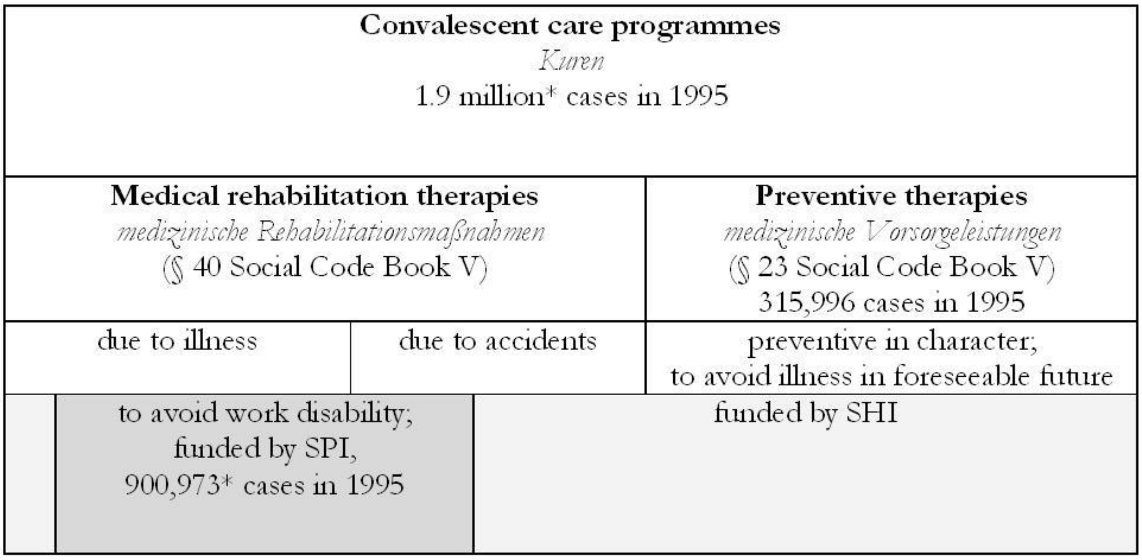
therapies at health spas. It may therefore come as no surprise that the German market for convalescent care is said to be the largest worldwide, at least when the booming wellness industry is not considered. In 1995, a total of €7.646 billion was spent on convalescent care, accounting for more than 4 percent of all health expenditures in Germany. The SHI spent €2.6 billion thereof and the Statutory Pension Insurance (SPI) spent €3.4 billion. Around 1,400 medical facilities with 100,000 full-time (equivalent) staff members treated 1.9 million patients, who stayed 31 days each on average (German Federal Statistical Office, 2010).

Both therapy forms, medical rehabilitation therapies and preventive therapies, require a physician's prescription, and the individual has to submit an application for treatment to his or her SHI sickness fund. The role of the patient in the application process is central. On the one hand, well informed patients may push their doctors to recommend them for convalescent care, and doctors may comply simply out of the fear of losing patients given the competition on the market and free choice of doctors by those insured under the SHI. On the other hand, patients may not accept their doctor's recommendation for convalescent care. After the application, the SHI fund determines whether the preconditions for treatment have been fulfilled and authorizes the therapy. The wording of the preconditions can be found in the German social legislation, Social Code Book V (SGB V, article 23 para. 1, article 40 para. 1). The legislation stipulates that for approval of preventive therapies, the patient must be suffering bad health to a degree that is likely to lead to an illness or disability in the foreseeable future. Hence, preventive therapies are not "purely" preventive, but provided to patients with early-stage health conditions that are likely to lead to fully fledged illnesses. Medical rehabilitation, by contrast, implies the diagnosis of an actual illnesss. After authorization by the SHI sickness fund, the prescribed treatment is provided in an approved medical facility under contract with the SHI fund. These medical facilities are usually located in scenic rural villages licensed by the state as *Kurorte*, or spa towns. For a village to be granted such a license, it needs to fulfill several conditions established in state legislation: very pure air, seaside location, or mineral springs. The idea of providing patients a healthy change of environment is integral to the treatment program.

Figure 4.1 gives an overview of the convalescent care programs available to people insured under the SHI. As has already been mentioned in the introduction, the umbrella term convalescent care programs includes two main types of treatment:

medical rehabilitation therapies and preventive therapies.[4]

Figure 4.1: Overview of Convalescent Care Programs for the SHI-Insured



Notes: author's illustration; figures are taken from the German Federal Statistical Office (2008);
* Figures include PHI insured

Medical rehabilitation therapies can be subdivided according to the medical reasons for the rehabilitation therapy. Medical rehabilitation therapy may be prescribed, first, to recover from an illness, or second, to recover from an accident. Medical rehabilitation therapies can also be classified by the funder of the therapy. The German SHI system follows the clear principle of "rehabilitation before pension." This states that the Statutory Pension Insurance (SPI) is legally obligated to pay for medical rehabilitation treatments that help to prevent permanent partial or total work disabilities. Thus, the SPI pays treatments for patients whose illness is severe enough to threaten their ability to work. This specific medical rehabilitation program

[4] German Social Law subdivides each of these two types into "inpatient" and "outpatient" treatment. "Inpatient" means that the therapy is carried out in a medical facility in a *Kurort* and therefore means that not only does the patient have to travel to a different place, he or she actually has to live on site at the facility. This terminology is confusing since *outpatient preventative therapies* also entail traveling to a *Kurort*, but in this case patients do not stay overnight at the facility but rather in a guesthouse of their own choice. For *inpatient preventive therapies* and *inpatient medical rehabilitation therapies* the SHI also covers room and board. The use of *outpatient medical rehabilitation therapies* was negligible in the 1990s and they are outside the scope of this chapter. These forms of treatment are provided in a medical facility in the patient's city of residence, which means that the patient can sleep at home during the program. Nowadays, these treatments have become increasingly prevalent.

– to avoid work disability – is at the same time the most important one in quantitative terms. Out of 1.9 million cases of convalescent care, 0.9 million were financed by the SPI to prevent work disability.

Preventive therapies are financed by the SHI. Alongside medical rehabilitation therapies, they constitute the second main type of convalescent care programs.[5] As can be inferred from Figure 4.1, preventive therapies accounted for about 300,000 cases in 1995.

The following program types can be identified by means of the SOEP data: First, the SOEP questionnaire asks whether the respondent has received any form of convalescent care in a medical facility in a spa town. Hence I can directly measure the overall incidence of convalescent care programs as visualized in Figure 4.1. Moreover, I can identify specific forms of medical rehabilitation. Second, medical rehabilitation therapies to recover from accidents at work can be identified, since the SOEP data includes a question on work accidents that required medical treatment. Third, I can directly identify medical rehabilitation therapies to avoid work disability since the SOEP includes a question on whether the cost of therapy was covered by the SPI or the SHI.

As I will explain in detail in the Data section, I generate three dependent variables for these directly measurable program types and use them in various econometric models to estimate the impact of the reform on these programs. Unfortunately, preventive therapies are not directly identifiable in the SOEP since no question refers to them specifically. Hence, how the increase in copayments affected preventive therapies cannot be assessed within a separate econometric model. However, as can be seen in Figure 4.1, knowledge about the reform effect for all convalescent care programs as well as medical rehabilitation therapies to avoid work disability can be used to roughly calculate the reform effect on preventive therapies. Under the assumption that the copayment doubling affected SHI-funded medical rehabilitation therapies in a similar way as SPI-funded medical rehabilitation therapies, it is possible to derive the reform effects on preventive therapies by means of a back-of-the envelope calculation. I apply such a back-of-the-envelope calculation to derive price elasticities for

[5] In Germany, there exist even more types of convalescent care programs. However, their overall importance is minor and they are outside the scope of this chapter. The purpose of "vocational rehabilitation" (*berufliche Rehabilitation*), for example, is to integrate disabled people into the labor market. Vocational rehabilitation was not affected by the reforms that I evaluate here. In 2007, only 68,000 cases were registered. These benefits are covered by unemployment insurance (UI) (Rauch et al., 2008).

preventive therapies. Details about the calculation are in Appendix D.

4.2.2 The Policy Reforms of the Convalescent Care System

At the end of 1996, the German government under Chancellor Kohl doubled the daily copayments for convalescent care treatments. This public health reform affected everyone insured under the SHI. The reason for the increase was the suspicion of a large degree of moral hazard in the market for convalescent care. Prior to the reform, experts estimated that around a quarter of all treatments prescribed were unnecessary (Schmitz, 1996; Sauga, 1996).

In West Germany as of January 1, 1997, copayments for medical rehabilitation therapies and inpatient preventive therapies were increased from DM 12 (€ 6.14) per day to DM 25 (€ 12.78) per day. In East Germany, the copayments were increased from DM 8 (€ 4.09) to DM 20 (€ 10.23) per day. This reflects an increase of 108 (150) percent.[6] To illustrate how drastic this copayment increase really was, I multiply the daily copayment rates by the average length of stay according to the Federal Statistical Office (German Federal Statistical Office, 2010). The absolute increase per treatment amounted to around € 150 in East and West Germany. Before the reform and in relation to the monthly net wages of those who received convalescent care in my sample, the total copayment sum per treatment was 12 percent of the net wage in East Germany and 13 percent in West Germany. After the copayment increase, the total copayment sum per treatment approximately doubled to 25 (East) and 24 (West) percent of the average monthly net wage.

Table 4.1 displays the various subgroups of insured people who were affected differently by the increase in copayments. Subgroups (5) to (8) were completely unaffected since they were insured under the PHI, which is unaffected by SHI copayment changes. Consequently, subgroups (5) to (8) later serve jointly as control group. By contrast, subgroups (1) to (4) – SHI-insured non-working people, public-sector

[6] Passed on November 1, 1996 this law is entitled *Gesetz zur Entlastung der Beiträge in der gesetzlichen Krankenversicherung (Beitragsentlastungsgesetz - BeitrEntlG), BGBl. I 1996 p. 1631-1633.* The time that needs to elapse between two treatment episodes in order for the insured person to become eligible again was extended from three to four years, and the standard length of both types of therapy was reduced from four to three weeks. Both changes were only effective conditional on the non-existence of urgent medical reasons for treatment. While the reduction of the regular length of stay is likely of negligible importance for the incidence of therapies, I will assess the impact of extending the waiting period in Section 4.5.

employees, apprentices[7], and self-employed respondents – were affected by the co-payment increase and constitute the treatment group. I generate a dummy variable called T which takes on the value one for subgroups (1) to (4) and zero for subgroups (5) to (8). The treatment indicator is then used in the regression models.

Table 4.1: Identification and Definition of Subgroups and Working Sample

	Affected by copayment increase
Non-working with SHI (1)	yes
Public-sector employees with SHI (2)	yes
Self-employed with SHI (3)	yes
Apprentices with SHI (4)	yes
Non-working with PHI (5)	no
Public-sector employees with PHI (6)	no
Self-employed with PHI (7)	no
Apprentices with PHI (8)	no

Ideally one may wish to compare each of these occupational groups separately, i.e., SHI vs. PHI non-working, SHI vs. PHI public-sector employees, SHI vs. PHI self-employed, and SHI vs. PHI apprentices. However, for some of these comparisons, the sample sizes are not large enough to obtain a precise estimate, especially not in the more refined specifications. For example, my sample includes only about 420 self-employed people and 450 apprentices per year who are insured under the SHI. In the robustness checks, however, I compare SHI and PHI non-working respondents and SHI and PHI public-sector employees separately.

Private sector employees who are insured under the SHI are not listed in Table 4.1.[8] In addition to the increase in copayments, they were affected by two other

[7] Although apprentices do not use the system as much as the other occupational groups, they do use it. The incidence of convalescent care treatment among them was 1.5 percent over all years.

[8] Private-sector employees who are insured under the PHI are also not included in my working sample. They were *only* affected by two additional reforms, which I explain in the following, but not by the increase in copayments. In my sample, they consist of only 150 respondents per year.

reforms: first, a new law allowed employers to deduct two days of paid vacation for every five days that an employee was unable to work due to convalescent care treatment. Second, this law, which became effective on October 1, 1996, also cut statutory sick pay for private-sector employees from 100 to 80 percent of foregone gross wages.[9] In the working paper version, I show that the net effect of all reform elements is of the same magnitude as the pure copayment effect and that the cut in paid vacation and sick pay had no significant effect on top of the copayment effect (Ziebarth and Karlsson, 2009). I have two explanations for this finding. First, the cut in sick pay did not necessarily pose a limitation on the insured since they might have faced a decision between going to rehabilitation or simply staying home to recover from the illness or accident. In any case, the patient would have been on sick leave. If necessary, physicians usually recommend treatments in spa towns, but if patients prefer to stay home on sick leave, their wishes are usually respected. Second, the cut in vacation days may not have been a binding constraint since many employees take all or part of their paid vacation to go to convalescent care in any case. Although entitled to take paid leave *in addition* to their paid vacation, many employees fear negative job consequences, especially when unemployment rates are high. Henceforth, I focus on the copayment effect.

4.3 Dataset and Variable Definitions

4.3.1 Dataset

The empirical analysis relies on microdata from the German Socio-Economic Panel Study (SOEP). The SOEP is an annual representative household survey that started in 1984 and sampled more than 20,000 persons in 2006. Wagner et al. (2007) provide further details. Information on convalescent care treatments is only available for two post-reform years. Hence, for the core analyses, I use data on the 1995 to 1999 waves, which include time-invariant information, current information, and retrospective information about the previous year. As the main dependent variables contain information about the calendar year prior to the interview, I employ data on

[9] There are no official numbers available on how many employees were de facto affected by the two reforms. As for the sick pay cut, union pressure and strikes forced many employers in various industrial branches to maintain 100 percent sick pay. Several sources suggest that the majority of the employees were effectively unaffected by the cut in statutory sick pay (Ridinger, 1997; Hans Böckler Stiftung, 2009).

the years 1994 to 1998.[10]

I exclude respondents under the age of 18, who are exempted from copayments, and focus on the subgroups that I have defined in Table 4.1. Missing values on the explanatory variables were imputed by single imputation and missing-value regressions.

4.3.2 Dependent Variables

The SOEP contains various questions about the health insurance and the usage of health services. In total, I generate three dependent variables on the incidence of convalescent care programs. The main dependent variable *Convalescent care* measures whether the respondent received convalescent care in a spa town in the year prior to the interview; it takes the value one if that was the case, and zero if not. In other words, *Convalescent care* measures the overall incidence of convalescent care programs as visualized in Figure 4.1. The variable has been generated from the following question, which was asked continuously from 1995 to 1999: "Were you admitted to preventive or medical rehabilitation treatment facility in a spa town in 199X?" In German, this question is even clearer because of the well-known umbrella term *Kur* and the inpatient treatment this entails at a different location from the recipient's place of residence, a *Kurort* or spa town, which minimizes measurement errors. The fact that we do not know the exact period of the therapy does not severely hamper the analysis, even more so given that such treatments are usually not carried out over Christmas or New Year's. Hence, there should be no doubt as to whether the therapy was in 1996 or in 1997.

By combining the main dependent variable *Convalescent care* with further questions, I generate two additional dependent variables that I employ to measure specific medical rehabilitation therapies. Both dependent variables constitute subsets of the main dependent variable *Convalescent care*. I call the first *Rehab to avoid work disability*. It takes the value one for respondents who have a one on their *Convalescent care* variable and who claimed that the SPI funded their treatment.[11] Since we know

[10] If the respondent was interviewed in two subsequent waves, e.g., in 1994 and 1995, I match the time-variant data from the first year dealing with the current (first) year with retrospective data from the second year dealing with the first year. For example, in 1995, respondents were asked about their current health status but about their insurance status during the previous year. Hence, I use the 1994 data on health status together with the 1995 data on insurance status if the respondent was interviewed in both years.

[11] The exact SOEP question reads: "Who paid the greater part of the costs? a.) Statutory

that the SPI only funds medical rehabilitation to prevent people from becoming unable to work, I thus capture employees who underwent such treatment (see Figure 4.1 and Section 4.2 for further details).

The second additional dependent variable is generated to measure the incidence of medical rehabilitation therapies for recovery from work accidents (see Figure 4.1). Respondents were asked whether they had been admitted to a hospital or whether they received medical treatment because of a work-related accident in the previous year.[12] Hence, the third dependent variable has a one for employees who had a work-related accident that required a hospital stay or medical treatment and whose *Convalescent care* variable takes on the value one. It is named *Rehab for recovery from accident*. I assume that the individuals underwent a medical rehabilitation therapy due to the work-related accident. Note that *Rehab for recovery from accident* is a subset of the main dependent variable *Convalescent care*; it is not necessarily mutually exclusive from *Rehab to avoid work disability*, since an employee might have had a work accident requiring medical rehabilitation that was also necessary to avoid work disability. However, in my sample and over the years 1994 to 1997, a work accident triggered only 3 percent of all medical rehabilitation therapies that were prescribed to avoid work disability.

Appendix C displays the summary statistics for all dependent variables.

4.3.3 Covariates

In the econometric specifications, I make use of various control variables. These control variables capture personal and family-related characteristics such as *age*, *female*, *immigrant*, *partner*, or *children*. Moreover, I control for educational characteristics by using data on the highest school degree obtained. An important determinant of the demand for convalescent care programs is the health status of the respondents, which I control for. I also include covariates that measure whether the person was employed full-time, part-time, marginally, or was non-employed.[13] I addition-

Pension Insurance b.) Statutory Health Insurance c.) other organization." As this question was only asked up to 1998, I cannot employ the 1999 wave in the models where this variable serves as a dependent variable.

[12] The exact SOEP question reads: "Were you in the hospital or did you receive medical treatment last year, in 199X, because of a work-related accident? a.) yes, received medical treatment b.) yes, went to the hospital c.) no"

[13] An anonymous referee has pointed out that especially the non-employment status may change quickly. Hence, the assignment of respondents to the treatment and control group might be impre-

ally control for gross monthly income and the equalized household income, which I obtain by dividing the household income by the root of all household members. To capture time-invariant regional characteristics, I make use of 15 state dummies. Regional labor market dynamics are controlled for by the inclusion of the annual state unemployment rate. Time trends are captured by year dummies. A list of the covariates, their means and standard deviations can be found in Appendix C.

4.4 Estimation Strategy

4.4.1 Difference-in-Differences

I would like to measure how the increase in copayments affected the incidence of convalescent care programs. Thinking of the policy intervention as a treatment, I fit a probit model of the form:

$$P[y=1|\boldsymbol{X}] = \Phi(\alpha + \beta\text{post97} + \gamma\text{T} + \delta\underbrace{(\text{post97*T})}_{\text{DiD}} + \boldsymbol{\psi'}\zeta) \tag{4.1}$$

where *post97* is a dummy that takes on the value one for post-reform years and zero for pre-reform years. T stands for the treatment indicator (see Section 4.2). The interaction term between both dummies gives us the difference-in-differences (DiD) estimator. To evaluate how the reform affected the outcome variable, henceforth, I always compute and display the marginal effect of the interaction term $\frac{\Delta\Phi(.)}{\Delta(\text{post97*T})}$.[14] $\Phi(.)$ is the cumulative distribution function for the standard normal distribution and $\boldsymbol{\psi'}$ is a vector including all personal, educational, and job-related controls as well as time dummies, state dummies, and the annual state unemployment rate.

cise. Since the SHI/PHI status is very stable over time, the imprecision lies between the different subgroups which are insured under the SHI as well as between the different subgroups which are insured under the PHI.

[14] Puhani (2008) has shown that the advice of Ai and Norton (2004) to compute the discrete double difference $\frac{\Delta^2\Phi(.)}{\Delta\text{post97}\Delta\text{T}}$ is not of relevance in nonlinear models when the interest lies in the estimation of a treatment effect in a difference-in-differences model. Using treatment dummies, the average treatment effect on the treated is given by $\frac{\Delta\Phi(.)}{\Delta(\text{post97*T})} = \Phi(\alpha + \beta\text{post97} + \gamma\text{T} + \delta\text{DiD} + \boldsymbol{\psi'}\zeta) - \Phi(\alpha + \beta\text{post97} + \gamma\text{T} + \boldsymbol{\psi'}\zeta)$ which is exactly what I calculate and present throughout the chapter.

4.4.2 Identification

The identification strategy relies on DiD estimation and hence on the assumption of a common time trend of the outcome variable for treated and controls in the absence of the policy intervention. This assumption should hold conditional on all available covariates. In almost all natural experiments and non-randomized settings, controlling for a rich set of covariates is important since control and treatment group differ with respect to most of the observed characteristics. This is also true in the present case, as Table 4.2 shows. In comparison to the control group, the treatment group includes more females and immigrants, and the employees are less educated. Moreover, the people in the control group are in worse health but younger than the treated. The controls are more likely to have a partner, children, and to be employed full-time.

As can be seen from Table 4.3, the most important driver of the demand for convalescent care programs is health status. Not surprisingly, age also plays a role, as well as income. Dropouts and immigrants are less likely to receive convalescent care, probably because of information asymmetries.

Again, I would like to stress that the econometric specifications adjust the sample composition to these various personal, educational, and job-related characteristics of the respondents. Recall that the health status of the respondents is observed and controlled for. Likewise, adjustments are made for time effects, persistent state differences, and the annual state unemployment rate.

The key identifying assumption, the common time trend assumption, is likely to hold. It assumes the absence of unobservables that generate different outcome *dynamics* for the treatment and control group. It is worth mentioning that a selection on observables story is plausible in the present setting. In the first place, it is the SHI/PHI insurance status that determines treatment (see Table 4.1). Almost all factors that determine whether respondents are insured under the SHI or PHI – such as occupational status and income – are observed.

A method to check the absence of distorting unobservable effects is to estimate placebo regressions for years without a reform. I will make use of this method in the next section.

Table 4.2: Variable Means by Treatment and Control Group

Variable	Treatment Group	Control Group
Convalescent Care	0.0445	0.0322
Personal characteristics		
Female	0.6140	0.3659
Age	46.9	44.9
Age squared	2,570	2,227
Immigrant	0.1724	0.0680
East Germany	0.3067	0.1314
Partner	0.6714	0.7483
Children	0.3646	0.4053
Good health	0.4647	0.6104
Bad health	0.2008	0.1099
Educational characteristics		
Dropout	0.0728	0.0271
Certificate after 8 years' schooling	0.4251	0.2089
Certificate after 10 years' schooling	0.2664	0.2977
Certificate after 12 years' schooling	0.0251	0.0486
Certificate after 13 years' schooling	0.1128	0.3834
Certificate degree	0.0780	0.0293
Job characteristics		
Full-time employed	0.1951	0.6702
Part-time employed	0.0463	0.0537
Marginally employed	0.0099	0.0078
Civil servant	0.0098	0.4245
Public servant	0.6780	0.6429
Self employed	0.0558	0.2597
Apprentice	0.0568	0.0119
Gross income per month	614	2,123
Equalized household income per month	1,294	1,916
Regional unemployment rate	12.3	11.0
N	38,589	4,375

In contrast to Appendix C, this table gives mean values separately for the treatment and control group. As detailed in Section 4.3, *Convalescent care* is the overall incidence of convalescent care programs (see also Figure 4.1).

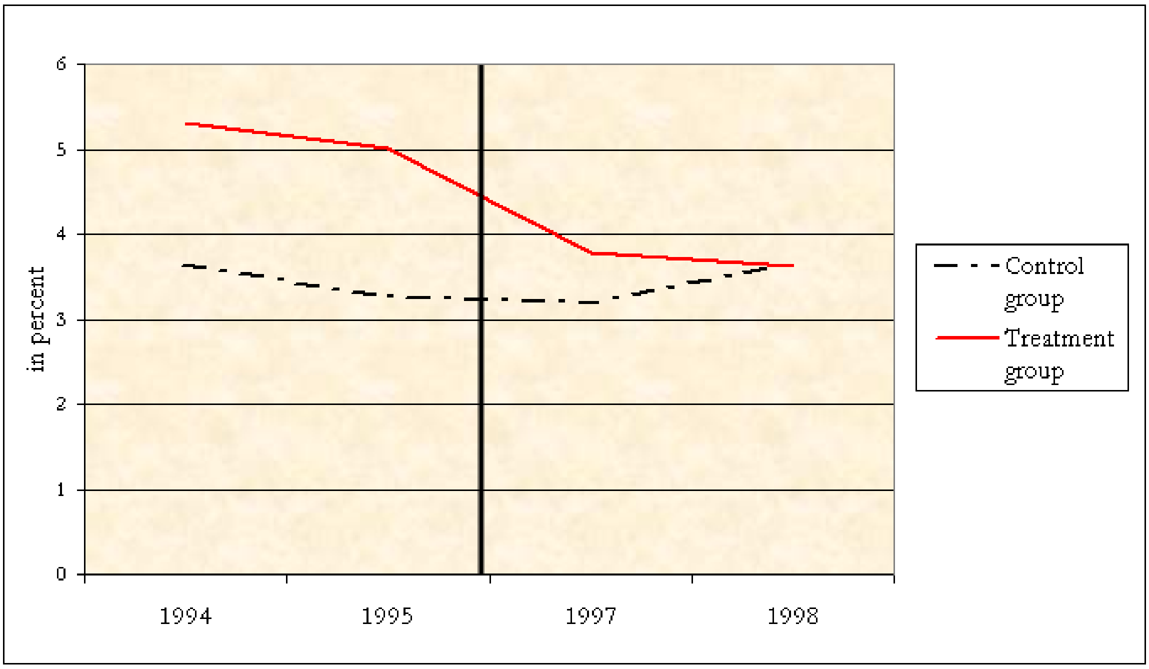
Table 4.3: Determinants of Convalescent Care Programs

Variable	Coefficient	Standard Error
Personal characteristics		
Female	-0.0012	0.002
Age	0.0022***	0.000
Age square/1,000	-0.0161***	0.003
Immigrant	-0.0105***	0.003
East Germany	0.0104	0.009
Partner	-0.0029	0.002
Children	-0.0002	0.002
Good health	-0.0225***	0.002
Bad health	0.0308***	0.002
Educational characteristics		
8 years of completed schooling	0.0111**	0.005
10 years of completed schooling	0.0208***	0.005
12 years of completed schooling	0.0211***	0.007
13 years of completed schooling	0.0149***	0.005
Other certificate	0.0116**	0.005
Job characteristics		
Full-time employed	-0.0017	0.004
Part-time employed	-0.0021	0.004
Marginally employed	0.0000	0.009
Gross wage per month/1,000	-0.0045***	0.001
Equalized household income/1,000	0.0031**	0.001
R-squared	0.0903	
χ^2	1,016	
N	42,964	

* p<0.10, ** p<0.05, *** p<0.01; marginal effects, which are calculated at the means of the covariates, are displayed. Dependent variable is *Convalescent care* and measures the incidence of all convalescent care programs. Standard errors in parentheses are adjusted for clustering on person identifiers. Regression includes state dummies. Left out reference categories are dropout and non-employed.

Figure 4.2 shows the evolution of the outcome variable for the treatment and control group over time.[15] Even without the correction for observables, we observe a parallel evolution in the two groups during the pre-reform years. After the reform, the incidence of convalescent care programs in the control group remained fairly stable, whereas we observe a clear and distinct decrease for the treatment group.

Figure 4.2: Incidence of Convalescent Care Programs by Treatment and Control Group



Compositional changes within the treatment and control group might have an impact on the outcome variable. For example, among the treatment group, the share of self-employed or public-sector employees may change over time which might affect or even produce the trend in the outcome variable. However, the share of the self-employed within the treatment group only fluctuated between 5.31 percent and 5.86 percent from 1994 to 1998. This is representative for the shares of the other subgroups, which are very stable over time. Additionally, in the empirical assessment, I contrast SHI and PHI non-working as well as SHI and PHI public-sector employees separately as robustness checks.

In recent years, there has been an extensive debate about the drawbacks and

[15] As will be shown later, there is evidence that distorting effects play a role due to the announcement of the reform at the end of 1995. Hence, the two uncontaminated pre-reform years, 1994 and 1995, are contrasted to the two post-reform years, 1997 and 1998.

limitations of DiD estimation. A particular concern is the underestimation of OLS standard errors due to serial correlation in the case of long time horizons and unobserved (treatment and control) group effects. To deal with the serial correlation issue, I focus on short time horizons. As Bertrand et al. (2004) have shown, the main reason for understating standard errors is rooted in serial correlation of the outcome and the intervention variable and is substantially alleviated when focusing on less than five periods. While there is consensus about the serial correlation problem, the issue with unobserved common group effects is a more controversial subject. If one takes the objection of Donald and Lang (2007) seriously, then it would not be possible to draw inferences from DiD analyses in the case of few groups, meaning that no empirical assessment could be performed. I subscribe to the view of Wooldridge (2006) who refers to it as (p. 18): *"DL [Donald and Lang] criticize Card and Krueger (1994) for comparing mean wage changes of fast-food workers across two states because Card and Krueger fail to account for the state effect (New Jersery or Pennsylvania) [...]. But the DL criticism in the $G = 2$ case is no different from a common question raised for any difference-in-differences analyses: How can we be sure that any observed difference in means is due entirely to the policy change? To characterize the problem as failing to account for an unobserved group effect is not necessarily helpful."*[16] Alongside the focus on short time spans to resolve serial correlation concerns, I use robust standard errors and correct for clustering at the individual level throughout the analysis.

A crucial issue in most studies that try to evaluate policy reforms is, besides the absence of a control group, selection into or out of the policy intervention. I can cope with concerns about selection since I am in the fortunate position of having a framework in which two almost totally independent health care systems exist side by side, as explained in Section 4.2. On the one hand, this provides a well defined control group. On the other hand, I do not need to fear that reform-induced selection

[16] In this very readable extended version of an older published AER paper (Wooldridge, 2003), Wooldridge (2006) discusses several other shortcomings and assumptions of the estimation approach proposed by Donald and Lang (2007). At another juncture, Wooldridge (2007) asks rhetorically whether introducing more than sampling error into DiD analyses was necessary or desirable. *"Should we conclude nothing can be learned in such settings?"*, he questions (p. 3). Moreover, he uses the well known Meyer et al. (1995) study, which is similar to the one at hand and obtains marginally significant results, as another example: *"It seems that, in this example, there is plenty of uncertainty in estimation, and one cannot obtain a tight estimate without a fairly large sample size. It is unclear what we gain by concluding that, because we are just identifying the parameters, we cannot perform inference in such cases. In this example, it is hard to argue that the uncertainty associated with choosing low earners within the same state and time period as the control group somehow swamps the sampling error in the sample means."* (p. 3 to 4).

has distorted the results, as there is virtually no switching between the SHI and the PHI, and since all SHI-insured persons are covered by universal health plans. Due to strict German regulations, a switch to the PHI was only legally allowed for a small fraction of optionally SHI-insured individuals, and I am able to identify and exclude these cases when running robustness checks. In another robustness check, I add a switching dummy to my model to see whether the results change.

Individuals insured under the SHI who were for some reason exempted from copayments are not identifiable. For example, people whose annual copayments for pharmaceuticals, health care services, or medical devices exceeded a certain percentage of their disposable household income could have applied for a case of hardship.[17] However, at that time, the *German Spa Association* claimed that the exemption clauses were widely unknown to the public. This should therefore not downwardly bias the results severely.

Furthermore, we need to consider the possibility of pull-forward effects. Convalescent care programs are usually planned several months or even years in advance. Certainly, preventive therapies are easier to schedule than medical rehabilitation treatments. Since the first policy reform plans were made public at the end of 1995 (Handelsblatt, 1995), it may be that a significant portion of the SHI-insured received their convalescent care therapy in 1996 instead of 1997. In the empirical application, I will check for anticipation effects.

Admittedly, it may have been that, due to rising awareness and increased political pressure, the SHI and SPI were more restrictive in their authorization of therapy programs during the period when the reforms were under political discussion, i.e., in 1996. As for anticipation effects that might have been triggered by the insured, one can test for such effects by either excluding the year 1996 from the analysis or by adding an interaction term between 1996 and the treatment indicator to the analysis.

To be able to fully attribute changes in the incidence to changes in the demand for convalescent care programs, supply-side effects should not play a role in this context. I have not found indications of supply-side constraints. In contrast, there have been reports about the deepest crisis in the market for convalescent care since the end of the Second World War (Handelsblatt, 1998). Dozens of medical facilities and health spas had to close and, hence, there is strong evidence that there was an

[17] The usual threshold is 2 percent of disposable household income; for people with chronic diseases it is 1 percent.

excess of supply. This is also supported by official statistics stating that the utilized bed capacity of all facilities strongly decreased, from 83.2 percent in 1996 to 62.3 percent in 1997 (German Federal Statistical Office, 2010).

Although representing average values for the whole of Germany, official data is also available on the average treatment length and the total number of days spent in inpatient medical facilities for convalescent care treatments. Since SOEP microdata are only available on the incidence of treatments, I might underestimate the reform's impact on the total number of treatment days consumed. However, for people insured under the SHI, the standard treatment length is codified in the Social Code Book and was reduced in the course of the reforms from four to three weeks. The exact treatment length is determined by medical personnel and the sickness fund and not by the patient. According to official data, the average treatment length for all insured individuals decreased from 31.0 (30.2) days in 1995 (1996) to 27.3 (26.4) days in 1997 (1998). The total number of convalescent care days consumed decreased from 57 million in 1995/1996 to 44.5 million in 1997/1998, representing a decrease of 22 percent (German Federal Statistical Office, 2010).

As has already been discussed in Section 4.2 and the Data section, the SOEP directly measures the overall incidence of convalescent care programs through one question. Likewise, I can directly identify medical rehabilitation therapies to avoid work disability since information is available about who funded the therapy and since we know that the SPI only covers this specific type of therapy (see Figure 4.1). As for medical rehabilitation therapies for recovery from work accidents, I need one assumption to identify this therapy form in the data: I plausibly assume that respondents who had a work-related accident requiring medical treatment and who were prescribed convalescent care therapy in the same year did in fact receive the therapy due to the accident. In this context, I have to point out that, in my sample, the number of respondents who belong to the control group and who underwent medical rehabilitation due to a work-related accident is so small that I cannot use the control group to identify the copayment effect. Hence, I discard the controls in this sub-specification. Since most work accidents happen to private-sector employees, I use private-sector employees who are insured under the SHI together with subgroups (2), (3), and (4) of Table 4.1 as my treatment sample to obtain a sufficiently large sample size.[18] This before-after estimator relies on the assumption that the demand

[18] As explained in Section 4.2, SHI private sector employees were also affected by two other reforms, but these had no effect on the demand for convalescent care which I show in the working

remained constant for the control group, which seems to be a reasonable assumption given Figure 4.2.

As a last point, I would like to emphasize that the identification strategy for the difference-in-differences regression models is based on various specifications. In total, I estimate models with three different dependent variables: *Convalescent care, Rehab to avoid work disability, Rehab for recovery from accident*. By this means, I automatically cross-check the consistency and plausibility of the identified reform effects.

4.5 Results

I have already examined Figure 4.2 in the previous subsection. It displays the unconditional incidence rates of convalescent care treatments by treatment and control group. While the incidence rate for the treatment group decreased sharply after the implementation of the policy reform, it remained fairly stable for the control group. Averaging over the pre- and post-reform years for both groups and taking the unconditional double difference yields a raw DiD estimate of -1.6 percentage points.

4.5.1 Copayment Effect on the Incidence of Convalescent Care Programs

Table 4.4 shows the effect of the increase in copayments on the overall incidence of convalescent care programs. Every column represents one model as in equation (4.1) and all models – except for the one in column (2) – estimate probit models. I always use an unbalanced panel. Since the coefficient of interest, displayed as DiD, is insensitive to the stepwise inclusion of sets of covariates, I present the full specifications and always display marginal effects.

Column (1) shows the effect when I estimate equation (4.1) and includes all covariates that are displayed in Appendix C. According to this "standard" model, I find that the increase in copayments has led to a decline in the incidence of convalescent care programs by 1.33 percentage points. The effect is significant at the 1.5

paper version. In the working paper, I also show by means of a graph that the incidence of convalescent care programs for private sector employees with SHI and the incidence for subgroups (2) to (5) run parallel (Ziebarth and Karlsson, 2009).

percent level. In column (2), the OLS model yields a highly significant effect of 1.6 percentage points.

Columns (3) and (4) test whether selection into or out of the treatment may have confounded the standard estimate in column (1). This is not the case since both estimates are very similar in size and significance to the estimate in column (1). In column (3), I exclude all respondents who were optionally insured under the SHI. They were the only population group that could have opted out of the SHI in response to the reform. All others were compulsorily insured under the SHI. In column (4), I add a switching dummy to the regression model. The switching dummy has a one for those who changed their SHI sickness fund or opted out of the SHI in 1996, 1997, or 1998. Switching between the SHI funds was allowed from January 1996 onwards to foster competition between the 600 SHI sickness funds that existed in 1996 (German Federal Statistical Office, 2010). Since the SHI sickness funds approve or deny applications for SHI-funded convalescent care treatments, it is imaginable that insured people switched to funds that had the reputation of being less restrictive than others in the authorization of such therapies. However, the copayment was doubled under federal law and applied unambiguously to all funds. I find that the copayment effect is exactly the same with or without the switching dummy.

Table 4.4: Copayment Effect on the Incidence of Convalescent Care Programs (I)

Variable	**Standard (1)**	**OLS (2)**	**w/o optional insured (3)**	**+ switching dummy (4)**		**(1) + DiD96 (5)**	**DiD96-98**	**+ income effects (6)**
DiD	**-0.0133****	**-0.0160*****	**-0.0123****	**-0.0133****	DiD	**-0.0103***	DiD98	**-0.0121***
	(0.0055)	**(0.0056)**	**(0.0056)**	**(0.0055)**		**(0.0058)**		**(0.0066)**
Treatment dummy	0.0056**	0.0088**	0.0056*	0.0055**	DiD96	**0.0113**	DiD97	**-0.0101****
(T)	(0.0028)	(0.0044)	(0.0029)	(0.0028)		**(0.0096)**		**(0.0046)**
Post-reform dummy	-0.0008	-0.0021	-0.0027	-0.0009	post1997	-0.0033	DiD96	0.0091
(*post1997*)	(0.0049)	(0.0062)	(0.0053)	(0.0049)		(0.0055)		(0.0084)
Dummy 1997	0.0011	0.0016	0.0003	0.0013	t1997	0.0011	DiD*gross-	**-0.0001**
(*t1997*)	(0.0021)	(0.0026)	(0.0022)	(0.0021)		(0.0022)	wage/1,000	**(0.0015)**
Dummy 1996	-0.0052***	-0.0083**	-0.0065***	-0.0049**	t1996	-0.0132**	DiD*equ.	**-0.0007**
(*t1996*)	(0.0019)	(0.0034)	(0.0021)	(0.0019)		(0.0059)	hhinc/1,000	**(0.0022)**
Dummy 1995	-0.0023	-0.0037	-0.0025	-0.0023	t1995	-0.0025	t1998	0.0004
(*t1995*)	(0.0019)	(0.0031)	(0.0019)	(0.0019)		(0.0019)		(0.0049)
Educational characteristics	yes	yes	yes	yes		yes		yes
Job characteristics	yes	yes	yes	yes		yes		yes
Personal characteristics	yes	yes	yes	yes		yes		yes
Regional unemployment rate	yes	yes	yes	yes		yes		yes
State dummies	yes	yes	yes	yes		yes		yes
Year dummies	yes	yes	yes	yes		yes		yes
R-squared	0.0910	0.0342	0.0915	0.0911		0.0911		0.0911
χ^2/F(.)	1,025	19.81	954	1,024		1,024		1,030
N	42,964	42,964	39,850	42,964		42,964		42,964

* p<0.1, ** p<0.05, *** p<0.01; marginal effects are displayed. They are calculated at the means of the covariates except for T(=1) and *DiD*(=1). Dependent variable is *Convalescent care* and measures the incidence of convalescent care programs (see Figure 4.1 and Section 4.3). Every column represents one regression model; all columns but (2) estimate probit models. The model in column (3) excludes all optional insured, the only group that is allowed to opt out of the SHI. The switching dummy in column (4) is -0.0074 (0.0061) and not significant at the 10 percent level. The model in column (5) includes an interaction term between T and *t1996* (DiD96). The model in column (6) is the most flexible specification and includes besides the DiD96 interaction also two interaction terms between T and *t1997* (DiD97) as well between T and *t1998* (DiD98). Thus the reform effect is not constraint to be the same over both post-reform years as in the other models. Standard errors in parentheses are adjusted for clustering on person identifiers.

Since the copayment doubling was first announced in December 1995, is it likely that the pre-reform year 1996 is contaminated by either pull-forward effects triggered by the insured or by supply-side effects triggered by SHI sickness funds or the SPI. The SHI and SPI might have been more restrictive in the authorization of treatments due to rising public awareness and political pressure. Including an interaction term between the treatment indicator and the year 1996 (DiD96) in column (5) does indeed yield some evidence that this might have been the case. Although the coefficient is imprecisely estimated, it has a positive sign – which points towards pull-forward effects of the insured – and causes the DiD coefficient to decrease slightly. However, the DiD coefficient is still significant at the 7.8 percent level and lies at -1.03 percentage points. Related to the 5 percent pre-reform incidence rate of convalescent care programs for the treatment group, this translates into a reform effect of 20.6 percent. In contrast to that, using the estimate of column (1), we would conclude that the demand for convalescent care programs decreased by 26 percent. Hence it seems reasonable to conclude that the copayment doubling has caused the demand for convalescent care programs to fall by 20 to 25 percent.[19]

Column (6) is the most flexible specification of Table 4.4. Instead of constraining the reform effect to be the same in both post-reform years, I allow the effect to differ between 1997 and 1998. At the same time, this is a test of the short-term effect of another reform element that was only briefly mentioned in Section 4.2: together with the increase in copayments, the waiting period for SHI-insured was increased from three to four years. The waiting period represents the time required to elapse between two treatments to become eligible again. However, the increase in the waiting period was only effective conditional on the nonexistence of urgent medical reasons for a therapy. As detailed in Section 4.2, people insured under the SHI have free choice of doctors and there are almost no waiting times for doctor appointments in Germany. Thus, it is unlikely that this change in the regulation had any substantial effect since finding a doctor who will issue a prescription for treatment is usually not difficult.

[19] This result also holds true when private sector employees who are insured under the SHI are included in the treatment group.

Table 4.5: Copayment Effect on the Incidence of Convalescent Care Programs (II)

Variable	**96 vs. 97 (1)**	**SHI funded (2)**	**West (3)**	**no health var. (4)**	**hospital < 7 nights (5)**	**hospital > 7 nights (6)**
DiD	**-0.0179***	**-0.0118****	**-0.0147*****	**-0.0142****	**-0.0079****	**-0.0062***
	(0.0098)	**(0.0053)**	**(0.0053)**	**(0.0060)**	**(0.0041)**	**(0.0034)**
Treatment dummy	0.0085**	0.0047***	0.0038	0.0078***	0.0039**	0.0012
(T)	(0.0034)	(0.0017)	(0.0027)	(0.0029)	(0.0019)	(0.0018)
Post-reform dummy			0.0049	-0.0024	-0.0022	0.0020
(*post1997*)			(0.0052)	(0.0056)	(0.0038)	(0.0026)
Dummy 1997	0.0077	0.0090*	0.0038	0.0013	-0.0016	0.0023
(*t1997*)	(0.0068)	(0.0047)	(0.0028)	(0.0023)	(0.0019)	(0.0013)
Dummy 1996		0.0013	-0.0021	-0.0058***	-0.0039***	-0.0013*
(*t1996*)		(0.0018)	(0.0029)	(0.0022)	(0.0015)	(0.0011)
Dummy 1995		-0.0003	-0.0037**	-0.0026	-0.0011	-0.0010
(*t1995*)		(0.0015)	(0.0037)	(0.0020)	(0.0014)	(0.0012)
Educational characteristics	yes	yes	yes	yes	yes	yes
Job characteristics	yes	yes	yes	yes	yes	yes
Personal characteristics	yes	yes	yes	yes	yes	yes
Regional unemployment rate	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes	yes	yes
R-squared	0.0941	0.0870	0.0947	0.0504	0.0672	0.1152
χ^2	482	658	704	544	518	718
N	16,935	33,560	30,555	42,964	42,186	41,924

* p<0.1, ** p<0.05, *** p<0.01; marginal effects are displayed; they are calculated at the means of the covariates except for T(=1) and DiD(=1). Dependent variable is *Convalescent care* and measures the incidence of convalescent care programs (see Figure 4.1 and Section 4.3). Every column represents one probit model. The model in column (1) contrasts 1997 to 1998 and thus estimates the short-term reform effect. The model in column (2) estimates the reform effect for convalescent care programs that are funded by the SHI (see Figure 4.1 for an overview). The model in column (3) focuses on West Germany. The model in column (4) excludes self-reported health measures. The models in columns (5) and (6) focus on respondents who received a convalescent care therapy and were either more than or less than 7 nights in hospital in the same year. Standard errors in parentheses are adjusted for clustering on person identifiers.

The increase in the waiting period forced patients who received a treatment in 1994 (1995) not to receive the next until 1998 (1999) instead of 1997 (1998) in the absence of medical reasons. Thus, if the increase in the waiting period had a substantial impact, I would measure a stronger reform effect for 1997 than for 1998. Column (6) shows that the reform effect in 1997 was even lower than in 1998. I take this as evidence that the increase in the waiting period had no significant (short-term) effect on the demand of convalescent care programs.

To test whether the reform effect differed by income, column (6) also includes interaction terms between the gross wage and DiD as well as between the equalized household income and DiD. We see that neither of the two coefficients is significantly different from zero and both are almost zero in size.

Column (1) of Table 4.5 contrasts the pre-reform year 1996 to the post-reform year 1997 and thus yields the short-run effect of the copayment increase, anticipation effects included. The estimate is at -0.018 percentage points and significantly different from zero.

Column (2) focuses on convalescent care programs that were funded by the SHI (see Figure 4.1 for an overview). The copayment effect is -0.012 percentage points and is significant at the 3 percent level. Compared to the pre-reform incidence rate, this is equivalent to a decrease of about 40 percent. This decrease, which is larger than the estimated 20 to 26 percent decrease for all convalescent care programs, makes sense since SHI-funded preventive therapies are likely to be much more price responsive than SPI-funded medical rehabilitation therapies to avoid work disability. According to the literature on health care demand, the price responsiveness is lower the more severe the illness and the more urgent the need for care.

The reform estimates for West Germany (column (3)) and the specification without health variables (column (4)) are quite similar in size to the "standard" estimate in column (1) of Table 4.4. I exclude health variables in column (4) since one might fear that the health status is endogenous if measured after a convalescent care therapy.[20]

The specifications in the last two columns of Table 4.5 focus on those who received

[20] Recall that health status refers to the time of the interview, whereas the information about convalescent care programs is sampled retrospectively for the previous calendar year. As explained at the beginning of Section 4.3, if a respondent was interviewed in two subsequent years, I match the current health status information in year t_0 with the convalescent care information of year t_1 which refers to year t_0. Since two-thirds of all interviews were carried out between January and March, the health status is likely to be measured before the medical treatment.

convalescent care therapy and were in a hospital either more than or less than seven nights in the same year. Both effects are significantly different from zero, although the effect for those with more than seven nights in hospital is slightly smaller in magnitude.

In the working paper version, I also show that balancing the sample as well as weighting the regressions with the inverse probability of not dropping out of the sample in the post-reform period does not change the results. I also present various models that exclude the year 1996, which is an alternative to the inclusion of DiD96 in columns (5) and (6) in Table 4.4. The results are almost identical (Ziebarth and Karlsson, 2009).

4.5.2 Copayment Effect on Convalescent Care: Refined Subgroup Comparisons

In the previous two tables, I have contrasted all subgroups that were affected by the increase in copayments jointly with all subgroups that were unaffected by the increase in copayments. In Table 4.6, I compare non-working respondents who were affected and unaffected by the reform as well as affected and unaffected public-sector employees. For both specifications, I provide probit and OLS estimates. As above, the OLS estimates are slightly larger in magnitude. However, all four models yield highly significant difference-in-differences estimates. When compared to the pre-reform program incidence, the decreases in demand are remarkably similar for both specifications. For the probit models they amount to around 35 percent. The finding that the these decreases in demand are larger than for all affected subgroups taken together might have many reasons. For example, it is likely that the overall demand of convalescent care programs includes a larger fraction of preventive therapies when non-working people and public-sector employees are considered.

Table 4.6: Copayment Effect on Convalescent Care Programs: Refined Subgroup Comparisons

	SHI non-working vs. PHI non-working		**SHI public-sectorvs. PHI public-sector**	
Variable	**Probit (1)**	**OLS (2)**	**Probit (1)**	**OLS (2)**
DiD97	-0.0213**	-0.0309**	-0.0158**	-0.0175**
	(0.0093)	(0.0141)	(0.0074)	(0.0081)
Educational characteristics	yes	yes	yes	yes
Job characteristics	yes	yes	yes	yes
Personal characteristics	yes	yes	yes	yes
Regional unempl. rate	yes	yes	yes	yes
State dummies	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes
R-squared	0.0855	0.0341	0.0853	0.0308
χ^2/F(.)	651	15.80	275	4.59
N	27,815	27,815	10,112	10,112

* p<0.1, ** p<0.05, *** p<0.01; in columns (1) and (3) marginal effects are displayed; they are calculated at the means of the covariates except for $T(=1)$ and $DiD(=1)$. Dependent variable is *Convalescent care* and measures the incidence of convalescent care programs (see Figure 4.1 and Section 4.3). Columns (1) and (2) contrast SHI non-working with PHI non-working, i.e., subgroup (1) with subgroup (5) (see Table 4.1). Columns (3) and (4) contrast SHI public-sector employees with PHI public-sector employees, i.e., subgroup (2) with subgroup (6). Columns (1) and (3) estimates probit models and columns (2) and (4) OLS models. Standard errors in parentheses are adjusted for clustering on person identifiers.

4.5.3 Copayment Effect on Medical Rehabilitation Therapies

Columns (1) and (2) of Table 4.7 estimate how the copayment doubling affected medical rehabilitation therapies for people who, without treatment, would run the risk of becoming unable to work. The SPI funds these programs, and I have information on who funded the therapy for the years 1994 to 1997. Hence, columns (1) and (2) in Table 4.7 use the dependent variable *Rehab to avoid work disability*. When compared with pre-reform incidence rates, the marginally significant probit estimate of -0.0025 translates into a reform decrease of 19.5 percent and the OLS effect would yield a decrease of about 24 percent. Note that the power of the statistical tests clearly decreases since I cannot use information on the convalescent care programs in 1998.

Table 4.7: Copayment Effect on the Incidence of Medical Rehabilitation Therapies

	Medical rehabilitation to avoid work disability		**Medical rehabilitation for recovery from accident**	
Variable	**Probit (1)**	**OLS (2)**	**Probit (3)**	**OLS (4)**
DiD97/post1997	-0.0025*	-0.0030**	-0.0012**	-0.0017**
	(0.0015)	(0.0014)	(0.0005)	(0.0008)
Educational characteristics	yes	yes	yes	yes
Job characteristics	yes	yes	yes	yes
Personal characteristics	yes	yes	yes	yes
Regional unempl. rate	yes	no	yes	yes
State dummies	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes
R-squared	0.1170	0.0128	0.069	0.003
χ^2/F(.)	431	8.98	114	1.92
N	33,024	33,024	34,339	34,339

* p<0.1, ** p<0.05, *** p<0.01; in columns (1) and (3), marginal effects are displayed; they are calculated at the means of the covariates except for $T(=1)$, *DiD97*(=1), and *post1997*(=1). Dependent variable in columns (1) and (2) is *Rehab to avoid work disability*, i.e., a dummy that captures the incidence of medical rehabilitation programs to prevent work disability and that are paid by the SPI (see Figure 4.1 and Section 4.3). Since no information about the funder of the treatment is available for 1998, *DiD97* identifies the copayment effect on the incidence. Dependent variable in columns (3) and (4) is *Rehab for recovery from accident*, i.e., a dummy that captures the incidence of medical rehabilitaion therapies due to a work accident (see Section 4.3 for details). Identification of the reform effects in columns (3) and (4) relies on a before-after estimator for subgroups (2), (3), (4) as well as all private sector employees who are insured under the SHI. Thus, the reform effect is identified by the post-reform dummy *post1997* (see subsection on Identification for details). Columns (1) and (3) estimates probit models and columns (2) and (4) OLS models. Standard errors in parentheses are adjusted for clustering on person identifiers.

Columns (3) and (4) of Table 4.7 estimate the reform effects on medical rehabilitation therapies for recovery from work accidents, i.e., employ the dependent variable *Rehab for recovery from accident*. In the subsection "Identification", I have already discussed that identifying the effect relies on a before-after estimator since too few respondents in the control group had work accidents, which is the reason for discarding the controls here. Moreover, I use all employees who are insured under the SHI as my treatment sample, i.e., subgroups (2), (3), and (4) of Table 4.1 plus all private-sector employees with SHI. Both Probit and OLS estimates have the same size and are significant at the 5 percent level. The probit estimate translates into a reform effect of 34 percent.

Despite having been estimated with smaller sample sizes, less precision, and under additional assumptions, it is remarkable that the reform estimates for medical rehabilitation therapies and the estimates for specific occupational groups fit the main estimates very well. In turn, the main estimates on the overall incidence of all convalescent care programs are strikingly robust to many alternative robustness checks, as Tables 4.4 and 4.5 demonstrate.

4.5.4 Placebo Reform Effects

As a final check to ultimately corroborate the common time trend assumption, I present placebo estimates for the years 1994 and 1995. That is, I pretend as if the reform had become effective in 1994 or 1995 and make use of information on the years 1993 to 1996. According to my approach above, I present placebo reform estimates on the overall incidence of convalescent care programs (column (1)) as well as on medical rehabilitation therapies to avoid work disability and on medical rehabilitation therapies for recovery from work accidents.

If any significant reform effects appeared, the assumption of a common time trend for treatment and control group in the absence of the policy intervention would be seriously challenged. However, as Table 4.8 demonstrates, this is not the case.

Table 4.8: Placebo Reform Estimates for 1994 and 1995

Variable	***Convalescent care***	***Medical rehababilitation*** *to avoid work disability*	***Medical rehababilitation*** *for recovery from accidents*
DiD95/t1995	-0.0011	0.0007	0.0004
	(0.0095)	(0.0026)	(0.0005)
DiD94/t1994	-0.0067	-0.0013	0.0009
	(0.0085)	(0.0023)	(0.0005)
Educational characteristics	yes	yes	yes
Job characteristics	yes	yes	yes
Personal characteristics	yes	yes	yes
Regional unempl. rate	yes	yes	yes
State dummies	yes	yes	yes
Year dummies	yes	yes	yes
R-squared	0.0534	0.0710	0.0445
χ^2	447	236	55
N	32,107	30,742	26,920

* p<0.1, ** p<0.05, *** p<0.01; marginal effects are displayed; they are calculated at the means of the covariates except for *DiD94*(=1), *DiD95*(=1), and *T*(=1) in columns (1) and (2). Every column estimates one probit model. In column (1), the dependent variable is *Convalescent care* and measures the incidence of convalescent care programs (see Figure 4.1 and Section 4.3). In column (2), the dependent variable is *Medical rehabilitation to avoid work disability*. In column (3), the dependent variable is *Medical rehabilitation for recovery from accidents*. In column (3), the reform effect is identified by a before-after estimator and thus the year dummies *t1994* and *t1995*. All models compare the same groups of (pseudo) treated and (pseudo) non-treated respondents as the non-placebo models. DiD95 is an interaction term between the treatment indicator and the year dummy *t1995*. DiD96 is an interaction term between the treatment indicator and the year dummy *t1996*. Standard errors in parentheses are adjusted for clustering on person identifiers.

4.5.5 Price Elasticities for Convalescent Care and Medical Rehabilitation Therapies

To calculate price elasticities, I make use of the regression results from the previous subsections. The formula for calculating arc price elasticities is:

$$\varepsilon_{q,p} = \frac{(q_1 - q_0)/\bar{q}}{(p_1 - p_0)/\bar{p}} \tag{4.2}$$

where q_1 represents the incidence of convalescent care programs for post-reform years and q_0 represents the incidence for pre-reform years. $\bar{q}$ is the average incidence rate over all years under consideration. Equivalently, $(p_1 - p_0)$ stands for the copayment increase and $\bar{p}$ is the average copayment rate over all years.

Thus, I plug the percentage point copayment reform effects for $(q_1 - q_0)$ into the formula and relate it to the average incidence rate over *all* years. For calculating the price elasticity of convalescent care programs in general, I take the estimate from column (1) of Table 4.4 and divide it by the overall incidence of convalescent care programs for the treatment group over the years 1994 to 1998, which was 4.45 percent. To calculate the price elasticity of medical rehabilitation therapies to avoid work disability, I take the estimate from column (1) of Table 4.7 and divide it by the overall incidence of this program type, which was 1.18 percent. Finally, to calculate the price elasticity of medical rehabilitation therapies for recovery from work accidents, I take the estimate from column (3) of Table 4.7, divide it by the overall incidence for all years and get a value of -41.5 (percent) for the numerator.

Concerning the denominator, I calculate the change in daily prices as $\frac{(25-12)}{(25+12)/2} = +70.27$ percent in West and $\frac{(20-8)}{(20+8)/2} = +85.71$ percent in East Germany and consider by simple weighting that 18.8 percent of all treatments were received by East Germans between 1994 and 1998 according to official statistics (German Statutory Pension Insurance, 2008).

The results are given in Table 4.9. The price elasticity of demand for convalescent care programs in general is estimated to be -0.41. The price elasticity of demand for medical rehabilitation therapies to avoid work disability is lower at -0.29. Finally, the price elasticity of demand for medical rehabilitation therapies for recovery from work accidents lies at -0.55. For the elasticity estimates in columns (1) and (3), I find that the 95 percent confidence interval does not include the zero, and for the elasticity estimate in column (2), the 90 percent confidence interval does not include the zero.

Table 4.9: Price Elasticity Estimates for Different Types of Convalescent Care Programs

All convalescent care therapies (1)	**Rehab to avoid work disability (2)**	**Rehab for recovery from work accidents (3)**	**Preventive therapies (4)**
-0.41	-0.29	-0.55	-1.08 up to -2.39
[-0.74;-0.08]	[-0.58;-0.02]	[-1.08;-0.10]	[-3.29;1.08] [-7.30;2.39]

Arc Price Elasticities are displayed and were calculated according to $\varepsilon_{q,p} = \frac{(q_1-q_0)/\bar{q}}{(p_1-p_0)/\bar{p}}$. Under consideration that copayments were increased by (25-12)/[(25+12)/2]= 70.27 percent in West and (20-8)/[(20+8)/2]= 85.71 percent in East Germany and by assuming that 18.8 percent of all therapies were undertaken by East Germans between 1994 and 1998 (German Federal Statistical Office, 2010). The estimated copayment effect from the regression models is plugged in for $(q_1 - q_0)$ in the formula above. To calculate the price elasticity of all convalescent care therapies, the DiD estimate from column (1) of Table 4.4 is used. To calculate the price elasticity of demand for medical rehabilitation therapies to avoid work disability, the DiD estimate from column (1) of Table 4.7 is taken, and for medical rehabilitation therapies for recovey from work accidents, the DiD estimate from column (3) of Table 4.7 is taken. The average incidence rates for the treatment group over all years are 0.0445, 0.0118, and 0.00299. Appendix D explains how the elasticities for preventive therapies are calculated. 95% confidence intervals, which are displayed in squared brackets, are calculated by means of the delta method; column (2) shows the 90% confidence interval.

As has already been mentioned several times, SOEP data do not contain explicit information on preventive therapy. Hence, there is no regression model that directly estimates the reform effect on the incidence of preventive therapy. However, since I know how both the demand for all convalescent care programs and the demand for SPI-funded medical rehabilitation therapies reacted, I can roughly assess the reform effect on the incidence for preventive therapies, as Figure 4.1 demonstrates. Once I know how the increase in copayments affected the demand for preventive therapies, I can also calculate elasticities. Appendix D gives details about this back-of-the-envelope calculation, the underlying assumptions, and how I calculated the upper and lower bound elasticity estimates.

The upper and lower bound estimates for the price elasticity of preventive therapies are -1.1 and -2.4, which means that the demand is price-elastic in contrast to medical rehabilitation therapies. This is in line with the literature on health care demand, which says that the price responsiveness for preventive care is higher than for other types of health care (Keeler et al., 1988; Zweifel and Manning, 2000). However, I am not aware of any concrete price elasticity estimate for preventive care with which to compare my estimates.

Note, however, that there is a high degree of uncertainty in the back-of-the-envelope calculation, such that the 95 percent confidence intervals for the price elasticity estimates of preventive therapies are so large that they include the zero. Strictly speaking, this would mean that the price elasticity for preventive therapies is not different from zero. On the other hand, as Appendix D explains, using official data on the post-reform decrease in demand for preventive therapies yields an elasticity estimate of -1.1 which reinforces the lower bound result from the back-of-the-envelope calculation.

4.6 Discussion and Conclusion

I have evaluated how a doubling of the copayment rate for convalescent care programs affected the demand for such programs in Germany. The main reason for implementing the reform was the suspicion of a large degree of moral hazard in the German market for convalescent care. Before the reform became effective, experts claimed that around a quarter of all convalescent care therapies were unnecessary (Schmitz, 1996; Sauga, 1996).

By means of a difference-in-differences approach, I estimate the causal effect of the increase in copayments on the incidence of convalescent care programs. The two-track German health care system allows me to rely on a well-defined control group – privately insured people who were entirely unaffected by the copayment increase for the publicly insured. Selection into or out of the treatment is not an issue in this setting due to strict legislative regulations that prevent switching between the health care systems.

My findings suggest that the copayment doubling has caused the overall demand for convalescent care programs to decrease by between 20 and 25 percent. The decrease in demand for medical rehabilitation therapies to avert total or partial work disabilities was a little bit lower than 20 percent, whereas the demand for medical rehabilitation therapies to recover from work accidents decreased by about 35 percent.

The estimated effects of the copayment increase allow me to calculate price elasticities of demand for various types of convalescent care programs. The price elasticity for medical rehabilitation therapies to avoid work disabilities amounts to around -0.3. The elasticity for medical rehabilitation therapies that help in recovering from work accidents lies around -0.5. In contrast to that, I find that the price elasticity for preventive therapies is elastic and is less than -1. To my knowledge, this is also the first attempt to calculate the price elasticity of preventive care. German preventive therapies that are covered by the SHI are a form of preventive care and educate people to adapt to a healthier lifestyle.

The question to what degree such policy reforms succeed in reducing moral hazard or whether they actually lead to adverse health outcomes is difficult to quantify and is beyond the scope of this chapter. The overall decrease in demand fits well with the a priori claims by health experts, who argued that a quarter of all therapies were unnecessary. Although it is unlikely that moral hazard was totally eliminated by the reforms, it is probable that the majority of the decrease is due to a reduction in moral hazard and led to greater efficiency in the convalescent care market. On the other hand, if medically necessary therapies were not provided, this may have led to adverse health outcomes. Especially in the case of preventive care, it is difficult to balance the prevailing degree of moral hazard against potential long-term health improvements that reduce health expenditures and exert positive external effects. Some studies have found positive health effects of health spa stays. Patients with chronic diseases experienced reductions in pain and blood pressure, and for a sample of employees, beneficial effects on physical and particularly mental health, such as improved sleep quality, were found (Sekine et al., 2006; Cimbiz et al., 2005; Constant et al., 1998). While two of these studies are purely correlation-based, Constant et al. (1998) estimate the short-term effects of a randomized trial on 224 patients with chronic lower back pain. However, I am not aware of studies that evaluate the long-term health effects of preventive therapies or medical rehabilitation therapies. Assessing the long-term effects of preventive care is a promising research field.

Appendix C

Table 4.10: Descriptive Statistics for the Working Sample

Variable	Mean	Std. Dev.	Min.	Max.	N
Dependent variables					
Convalescent care	0.0433	0.2035	0	1	42,964
Rehab to avoid work disability	0.0138	0.1167	0	1	33,024
Rehab for recovery from accident	0.0030	0.0547	0	1	34,339
Covariates					
Personal characteristics					
Female	0.5887	0.4921	0	1	42,964
Age	46.8	18.6	18	99	42,964
Age squared	2,535	1,843	324	9,801	42,964
Immigrant	0.1618	0.368	0	1	42,964
East Germany	0.2888	0.4532	0	1	42,964
Partner	0.6792	0.4668	0	1	42,964
Children	0.3688	0.4825	0	1	42,964
Good health	0.4796	0.4992	0	1	42,964
Bad health	0.1916	0.3931	0	1	42,964
Educational characteristics					
Drop out	0.0687	0.2499	0	1	42,964
8 years of completed schooling	0.4031	0.487	0	1	42,964
10 years of completed schooling	0.2696	0.4405	0	1	42,964
12 years of completed schooling	0.0275	0.1621	0	1	42,964
13 years of completed schooling	0.1403	0.345	0	1	42,964
Other certificate	0.0731	0.2575	0	1	42,964
Job characteristics					
Full-time employed	0.2435	0.4292	0	1	42,964
Part-time employed	0.047	0.2117	0	1	42,964
Marginally employed	0.0097	0.0978	0	1	42,964
Gross wage per month	768	1,272	0	5,1129	42,964
Equivalized household income	1,357	681	21	1,6269	42,964
Regional unemployment rate	12.2	3.9	7	21.7	42,964

Appendix D

Calculating Price Elasticities for Preventive Therapies

From the regression models, I know how the increase in copayments affected the overall demand for convalescent care programs. In addition, I know how the copayment increase affected the demand for medical rehabilitation programs to avoid work disability. Since I also know that 15.13 percent of all convalescent care programs are preventive therapies (German Federal Statistical Office, 2010), it is possible to approximate how the increase in copayments affected the demand for preventive therapies. This can be inferred from Figure 4.1. However, one assumption is needed: namely, that the demand for SHI-funded medical rehabilitation therapies reacted in the same way as SPI-funded medical rehabilitation therapies. One can formulate:

$$\frac{q_1^c - q_0^c}{\bar{q}^c} = \rho_r \frac{q_1^m - q_0^m}{\bar{q}^m} + \rho_p \frac{q_1^p - q_0^p}{\bar{q}^p} \tag{4.3}$$

where $q_1^c - q_0^c$ is the copayment induced percentage point decrease of all convalescent care programs and $\bar{q}^c$ is the incidence rate over all years with q_1 being the incidence rate in the post-reform years and q_0 being the incidence rate in the pre-reform years. The superscript m stands for SPI funded medical rehabilitation programs and p stands for preventive therapies. ρ_r is the ratio of SPI- *and* SHI-funded medical rehabilitation programs, which is 0.8487 according to the German Federal Statistical Office (2010) and ρ_p is the ratio of preventive therapies, which is 0.1513. Plugging in and solving for $\Delta q^p = \frac{q_1^p - q_0^p}{\bar{q}^p}$ yields that the copayment increase has led to a decrease in the demand for preventive therapies by 78.7 percent.

When I divide the 78.7 percent decrease in the demand for preventive therapies by the formula for the copayment increase from equation (4.2), $\Delta p = \frac{p_1 - p_0}{\bar{p}} = 0.7317$, I obtain an elasticity of -1.08. However, I consider this as an lower bound estimate since 55 percent of all preventive therapies were *outpatient* preventive therapies at that time (German Federal Statistical Office, 2010). "Outpatient" means that patients do not stay overnight at the medical facility but rather in a guesthouse or hotel of their own choice. Copayments are not charged for outpatient preventive therapies but patients have to pay for food and overnight expenses on their own. I want to calculate the price elasticity for *inpatient* preventive therapies, in which case patients stay overnight in the medical facility. For inpatient preventive therapies, copayments do exist and were doubled as described in Section 4.2. Although the SOEP question on convalescent care programs asks explicitly about "inpatient treatment facilities", respondents might have misunderstood the question since outpatient preventive therapies also entail traveling to a spa town. When I divide the derived 78.7 percent decrease in demand for health spa treatments by $0.45 * \Delta p$, I get the upper bound elasticity of -2.39.

Interestingly, official data are also surprisingly in line with my back-of-the-envelope estimate on the decline in demand for preventive therapies. Using data from the German Federal Statistical Office (2010) on the number of inpatient preventive therapies for SHI insured for which copayments were charged, Δq^p amounts to 81.2 percent. This is remarkably close to my roughly estimated 78.7 percent. When dividing this decrease by the average increase in copayments Δp, I end up with an elasticity estimate that is derived from official data and that is -1.1. It lies between my upper and lower bound elasticity estimate from above, though much closer to the lower bound estimate. Since I have calculated the upper bound under the assumption that many respondents have misunderstood the SOEP question, the accuracy of the back-of-the-envelope calculation is at the same time indirect evidence that the SOEP question captured very well all medical rehabilitation and inpatient preventive therapies; it is also indirect evidence for the accuracy of my reform effect estimate on medical rehabilitation therapies to avoid work disabilities, i.e., $\frac{q_1^m - q_0^m}{\bar{q}^m}$ in equation 4.3 above.

In contrast to the elasticities for medical rehabilitation therapies, the elasticities for preventive therapies are likely to be larger than -1, which means price-elastic. This is at the same time the first price elasticity estimate for preventive care I am aware of.

Chapter 5

Assessing the Effectiveness of Health Care Cost Containment Measures

Abstract

Rising health expenditures are an issue of increasing concern in the industrialized countries. This chapter is the first to empirically evaluate the effectiveness of four different health care cost containment measures within an integrated framework. The four measures investigated were introduced in Germany in 1997 to reduce moral hazard and public health expenditures in the market for convalescent care. Various subpopulations were affected by these reforms in different ways. Using SOEP panel data and difference-in-differences methods, I assess the causal reform effects of these cost-containment measures on the demand for convalescent care. Doubling the daily copayments was clearly the most effective cost containment measure, resulting in a reduction in demand of about 20 percent. Indirect measures such as allowing employers to cut statutory sick pay or paid vacation during health spa stays did not significantly reduce demand.

5.1 Introduction

For decades health expenditures have been increasing exponentially in almost all of the industrialized countries. In the US, health spending increased a staggering 787 percent from 1980 to 2007. In reunified Germany, health expenditures increased from 1992 to 2008 by 60 percent, today consuming more than 10 percent of GDP (German Federal Statistical Office, 2010). In light of these figures, it is no surprise that rising health care costs are one of the most contentious issues of public debate at present and a matter of great concern for policy makers worldwide.

Researchers have identified various key factors behind rising health expenditures, including demographic change, increasing national incomes, and technological change. Joseph P. Newhouse was the first to identify technological change as the dominant driving force, a conjecture that is difficult to prove empirically (Newhouse, 1992; Okunade, 2004; Di Matteo, 2005; Civan and Koksal, 2010).

While the main causes of rising health expenditures seem clear, the question of how to deal with them remains unresolved. There is an extremely wide variety of organizing health care systems in different countries, but none of them has clearly emerged as the optimal model. This comes as no surprise if one thinks about the very different objectives that the various health care systems are designed to achieve: reducing the burden on the social security system and taxpayers, achieving equal access to care, providing universal coverage, avoiding state rationing, allowing freedom to choose medical providers and insurance plans, or promoting medical progress, to name just a few.

The literature has analyzed the optimal organization of health care theoretically as well as empirically, although the majority of work has been theoretical in nature. Some attention has been given to the supply side, particularly to the question of how to optimally organize and finance a hospital system with the aim of balancing quality of care against costs (Ellis and McGuire, 1996; Sloan et al., 2001; Propper et al., 2004; Bazzoli et al., 2008). Analogously, the same question can be raised for the outpatient sector and physicians (Mariñoso and Jelovac, 2003; Dusheiko et al., 2006; Karlsson, 2007). Especially in the US—a market that is still dominated by private health care providers—there is considerable debate surrounding the question of whether Health Maintenance Organizations (HMOs) can help reduce health expenditures while maintaining quality (Goldman et al., 1995; Hill and Wolfe, 1997; Keeler et al., 1998; Deb and Trivedi, 2009). In Europe, on the other hand, key concerns revolve around issues of direct rationing (by public authorities) and indirect rationing (through waiting times) (Propper et al., 2002; Schut and de Ven, 2005; Felder, 2008; Siciliani et al., 2009).

In the demand-side research, cost-sharing has been identified as the main tool used to reduce moral hazard and overconsumption of medical services (Pauly and Blavin, 2008; van Kleef et al., 2009). In this strand of the literature, the RAND Health Insurance Experiment (HIE) is still the largest and most influential health policy study to this day. In this study from the 1970s, families at six different sites in the US were randomly assigned to 14 different health insurance plans with a varying degree of cost-sharing and observed for periods up to five years (Manning et al., 1987). Since then, a great amount of publications on the impact of cost-sharing on the demand for medical care have emerged from the HIE, most of them from the 1980s (see Zweifel and Manning (2000) for an overview). But outside the US and its private health insurance system, there is only scanty empirical evidence of causal effects of cost-containment measures on the demand for health care. A handful of studies have empirically investigated how increased copayments affect the demand for doctor visits (Chiappori et al., 1998; Voorde et al., 2001; Cockx and Brasseur, 2003; Winkelmann, 2004; Gerfin and Schellhorn, 2006). Schreyögg and Grabka (2010) analyzed the effects of the copayments for doctor visits introduced in Germany. Using a difference-in-differences setup similar to the one in this chapter, as well as the same dataset, they did not find any significant behavioral reactions in the aftermath of the reform.

To the best of my knowledge, this is the first study to evaluate the effectiveness of four different cost containment measures within an integrated framework. In Germany, from 1997 on, various health reforms were implemented to reduce the demand for convalescent care. Before the reforms went into effect, experts claimed that around a quarter of all convalescent care therapies were unnecessary (Schmitz, 1996; Sauga, 1996). In 1995, 1.9 million patients in Germany underwent convalescent care therapy and more than €7 billion (0.4 percent of GDP) was spent on these programs (German Federal Statistical Office, 2010).

The first reform doubled the daily copayments for convalescent care. The second increased waiting times between two treatments and reduced the legally codified standard length of the therapy. The third reform gave employers the right to deduct two days of paid vacation for every five days that employees were unable to work because of being in convalescent care. The fourth reform cut statutory sick pay from 100 to 80 percent of foregone gross wages during convalescent care: employees are entitled statutory sick pay when absent from work due to convalescent care.

The first two reforms only affected people insured under the German Statutory Health Insurance (SHI), while people insured under the second tier of the German health insurance system—the Private Health Insurance (PHI)—were not affected. The other two reforms, which concerned the cut in paid leave, only affected private-sector employees. Thus, I can define various subgroups that were affected differently by the reforms. By means of conventional difference-in-differences models and SOEP panel data, I then disentangle the causal effects of these cost containment measures on the demand for convalescent care.

My empirical results show that the copayment doubling was, by far, the most effective cost containment instrument. It led to a significant decrease in the demand for convalescent care programs of about 20 percent. Moreover, descriptive evidence from administrative data suggests that the reduction in the legally defined standard length of the therapies was effective in reducing the average duration of treatments. However, I do not find evidence that the cuts in paid leave reduced the demand for convalescent care programs. Based on administrative data, back-of-the-envelope calculations suggest that all reforms jointly reduced annual public spending for convalescent care by €800 million or 13 percent. Although the length of treatments decreased, the doubling of daily copayments raised additional revenues of about €400 million per year.

In the next section, I describe some features of the German health care system and give more details about the reform. In Section 5.3, the dataset and the variables used are explained, and in the subsequent section, I specify my estimation and identification strategy. Estimation results are presented in Section 5.5 and I conclude with Section 5.6.

5.2 The German Health Care System and the Policy Reforms

The German health care system is actually comprised of two independent health care systems that exist side by side. The more important of the two is the Statutory

Health Insurance (SHI), which covers about 90 percent of the German population. Employees whose gross income from salary is below a defined income threshold (2009: €4,050 per month) are compulsorily insured under the SHI. High-income earners who exceed that threshold as well as self-employed people have the right to choose between the SHI and private health insurance. Non-working spouses and dependent children are covered at no cost by the SHI family insurance. Special regulations apply to particular groups such as students and the unemployed, but most of these are SHI-insured. Everyone insured under the SHI is subject to a universal benefit package, which is determined at the federal level and codified in the Social Code Book V (SGB V). Coinsurance rates[1] are prohibited in the SHI and thus, apart from copayments, health services are fully covered. The SHI is one pillar of the German social security system (German Ministry of Health, 2008).

The SHI is primarily financed by mandatory payroll deductions that are not risk-related. For people in gainful employment, these contributions are split equally between employer and employee up to a contribution ceiling (2009: €3,675 per month). Despite several health care reforms that have tried to remedy the problem of rising health care expenditures, contribution rates rose from 12.6 percent in 1990 to 14.9 percent in 2009, mainly due to demographic changes, medical progress, and system inefficiencies (German Federal Statistical Office, 2010).

The second track of the German health care system is Private Health Insurance (PHI). The main groups of private insurance holders are private-sector employees above the aforementioned income threshold, public-sector employees[2], and the self-employed. Privately insured people pay risk-related insurance premiums determined by an initial health checkup. The premiums exceed the expected expenditures in younger age brackets, since health insurance providers build up reserves for rising

[1] Coinsurance rates are important for private health insurance providers. They differ from copayments. While a copayment is typically a fixed amount that the insured person has to pay per day of treatment or for specific medical devices or medications, a coinsurance rate defines a percentage of the costs that an insured person has to pay when using the system. For example, private health insurance providers may offer 80/20 health plans in which the insured person pays 20 percent of all costs incurred while the health insurance provider pays the remaining 80 percent. Often, health insurance providers limit the total amount that an individual has to spend out-of-pocket with a so-called coinsurance cap, which might be €2,000 per year.

[2] We need to distinguish between two types of employees in the German public sector: first, civil servants with tenure (*Beamte*), henceforth called "civil servants," most of whom purchase PHI to cover the 50 percent of health expenditures that the state does not reimburse (*Beihilfe*), and second, employees in the public sector without legal tenure (*Angestellte im öffentlichen Dienst*), henceforth called "public servants," who receive some additional benefits but are mainly insured under the SHI (under the same conditions as everyone else)

expenditures with increased age. Coverage is provided under a range of different health plans, and insurance contracts are subject to private law. Consequently, in Germany, public health care reforms apply only to the SHI, not to the PHI.

It is important to keep in mind that compulsorily insured persons have no right to choose the health insurance system or benefit package. They are compulsorily insured under the standard SHI insurance scheme. Once an optionally insured person (a high-income earner, self-employed person, or civil servant) opts out of the SHI system, it is practically impossible to switch back. Employees above the income threshold are legally prohibited from doing so, while those who fall below the income threshold in subsequent years may do so under certain conditions, but any reserves that they have built up under PHI policies are not portable (neither between PHI and SHI, nor between the different private health insurance providers).[3] In reality, switching to a private health insurance provider may be regarded as a lifetime decision, and switching between the SHI system and PHI—as well as between PHI providers—is therefore very rare.

5.2.1 The German Market for Convalescent Care

In Europe, and especially in Germany, there is a long tradition of health spa treatments to improve poor health. Since the time of the Roman Empire, doctors have been sending patients to "take the waters" to recover from various disorders. In Germany, convalescent care treatments are usually combined with various types of physical therapy, often including electrotherapy, massage, underwater exercise, ultrasonic therapy, health and diet education, stress reduction therapy, and cold and hot baths as well as mud packs. Convalescent care therapies require the patients to follow a strict daily schedule.

The German SHI is one of the few health insurance systems worldwide that, apart from small copayments, fully covers convalescent care therapies at health spas. It may therefore come as no surprise that the German market for convalescent care is said to be the largest worldwide, at least when the booming wellness industry is not considered. In 1995, a total of €7.646 billion was spent on convalescent care, accounting for more than 4 percent of all health expenditures in Germany. Around 1,400 medical facilities with 100,000 full-time (equivalent) staff members treated 1.9

[3] Until 2009, accrued reserves for rising health expenditures with increased age were not portable. But since January 1, 2009, a strictly defined level of portability between PHI providers is compulsory.

million patients, who stayed 31 days each on average (German Federal Statistical Office, 2010).

Convalescent care therapy—referred to in Germany as a *Kur* or cure—requires a physician's prescription, and the individual has to submit an application for treatment to his or her SHI sickness fund. The role of the patient in the application process is central. On the one hand, well informed patients may push their doctors to recommend them for convalescent care, and doctors may comply simply out of the fear of losing patients given the competition on the market and free choice of doctors for those insured under the SHI. On the other hand, patients may not accept their doctor's recommendation for convalescent care. After the application, the SHI fund determines whether the preconditions for treatment have been fulfilled and authorizes the therapy. The wording of the preconditions can be found in the German social legislation, Social Code Book V (SGB V, article 23 para. 1, article 40, para. 1). After authorization by the SHI sickness fund, the prescribed treatment is provided in an approved medical facility under contract with the SHI fund. These medical facilities are usually located in scenic rural villages licensed by the state as *Kurorte*, or spa towns. For a village to be granted such a license, it needs to fulfill several conditions established in state legislation: pure air and location near the seaside or mineral springs. The idea of providing patients a healthy change of environment is integral to the treatment program.

5.2.2 The Cost Containment Policy Reforms

At the end of 1996, the German government under Chancellor Kohl implemented four health care reforms. The first three of these were designed to dampen the demand for convalescent care programs, based on the suspicion of a high degree of moral hazard in the market for convalescent care. Prior to the reform, experts estimated that around a quarter of all treatments prescribed were unnecessary (Schmitz, 1996; Sauga, 1996). The fourth reform was designed to tackle moral hazard in the decision to take sick leave and may have indirectly affected the demand for convalescent care as well.

The first reform doubled daily copayments. In West Germany, as of January 1, 1997, copayments for convalescent care therapies were increased from DM 12 (€ 6.14) per day to DM 25 (€ 12.78) per day. In East Germany, the copayments were increased from DM 8 (€ 4.09) to DM 20 (€ 10.23) per day. This reflects an increase of 108 (150) percent.[4] To illustrate how drastic this copayment increase really was, I multiply the daily copayment rates by the average length of stay according to the Federal Statistical Office (German Federal Statistical Office, 2010). The absolute increase per treatment amounted to around €150 in East and West Germany. Before the reform and in relation to the monthly net wages of those who received convalescent care in my sample, the total copayment per treatment was 12 percent of the net wage in East Germany and 13 percent in West Germany. After the copayment increase, the total copayment sum per treatment approximately doubled to 25 (East) and 24 (West) percent of the average monthly net wage.

The second reform reduced the standard length of convalescent care therapies from four to three weeks. Only the medical personnel of the facility—after consultation with the sickness fund—have the authority to approve deviations from the standard length of therapy, which is codified by law. Together with this reduction in therapy duration, waiting times were increased from three to four years between treatments. Both reform elements—the reduced standard length of therapy and the extended waiting period—are only effective conditional on the non-existence of urgent medical reasons for treatment.

The third reform allowed employers to deduct two days of paid vacation for every five days that an employee was unable to work due to convalescent care therapy. The fourth reform decreased statutory short-term sick pay from 100 to 80 percent of foregone gross wages. German social legislation provides employees with paid leave for convalescent care treatments *in addition* to paid vacation. Hence, one would expect that the latter two reforms, which allowed employers more leeway in reducing paid leave, had an effect on the demand for convalescent care.[5]

Table 5.1 displays the various subgroups of insured people who were affected differently by the four cost containment measures. Subgroup (1) comprises the vast majority of Germans: private-sector employees who are insured under the SHI. They were affected by all reforms discussed above. I define them as *Treatment Group 1*.

[4] Passed on November 1, 1996 this law is the *Gesetz zur Entlastung der Beiträge in der gesetzlichen Krankenversicherung (Beitragsentlastungsgesetz - BeitrEntlG), BGBl. I 1996 p. 1631-1633.*

[5] Passed on September 15, 1996, this law is the *Arbeitsrechtliches Gesetz zur Förderung von Wachstum und Beschäftigung (Arbeitsrechtliches Beschäftigungsförderungsgesetz), BGBl. I 1996 p. 1476-1479.* The law went into effect on October 1, 1996.

Table 5.1: Identification and Definition of Subgroups and Subsamples

	Reform 1: Copayment doubling	**Reform 2: Waiting time increase**	**Reform 3: Paid vacation reduction**	**Reform 4: Sick pay decrease**
Private sector with SHI (1) (*Treatment Group 1*)	yes	yes	yes	yes
Self-employed with SHI (2)	yes	yes	no	no
Non-working with SHI (3)	yes	yes	no	no
Public sector with SHI (4)	yes	yes	no	no
Apprentices with SHI (5) (*Treatment Group 2*)	yes	yes	no	no
Self-employed with PHI (6)	no	no	no	no
Non-working with PHI (7)	no	no	no	no
Public sector with PHI (8)	no	no	no	no
Apprentices with PHI (9) (*Control Group*)	no	no	no	no

In contrast, subgroups (2) to (5) were not affected by either the cut in statutory sick pay or the cut in paid vacation. Non-working and self-employed people are not eligible for paid leave. Public-sector employees and apprentices were exempted from the cuts in paid leave for political reasons. However, since they were insured under the SHI, they were affected by the first two reforms. I call these subgroups jointly *Treatment Group 2*.

Subgroups (5) to (9) were completely unaffected by all legislative changes; they also consist of the non-working, the self-employed, apprentices, and public sector employees, none of whom were affected by the cut in paid leave. However, in contrast to *Treatment Group 2*, subgroups (5) to (9) were insured under the PHI and, thus, reforms one and two did not apply to them either. I define subgroups (5) to (9) jointly as *Control Group*.[6]

[6] Private-sector employees who are insured under the PHI are not included in my working sample. They were *only* affected by reforms three and four but not by the increase in copayments or in waiting times. However, in my sample, they consist of only 150 respondents per year and, thus, I cannot use them to obtain precise estimates.

In total, I obtain three mutually exclusive subsamples that were affected differently by the reforms. Thus, in the empirical assessment, I use three distinct main models in which I compare these subsamples to evaluate the effectiveness of the reforms. To this end, I generate three treatment indicators that I will explain in more detail in Section 5.3.3 below.

5.3 Dataset and Variable Definitions

5.3.1 Dataset

The empirical analysis relies on micro data from the German Socio-Economic Panel Study (SOEP). The SOEP is an annual representative household survey that started in 1984 and meanwhile includes more than 20,000 respondents. Wagner et al. (2007) provide further details. Information on convalescent care treatments is only available for two post-reform years. Hence, for the core analyses, I use data on the 1995 to 1999 waves, which include time-invariant information, current information, and retrospective information about the previous year. Since the main dependent variables contain information about the calendar year prior to the interview, I employ data on the years 1994 to 1998.[7]

I exclude respondents under the age of 18, who are exempted from copayments, and focus on the subgroups that I have defined in Table 5.1.

5.3.2 Dependent Variable and Covariates

The SOEP contains various questions about health insurance and the use of health care services. The dependent variable *convalescent care* measures whether a respondent received convalescent care at a health spa in the calendar year prior to the interview; it takes the value one if that was the case, and zero if not. In other words, *convalescent care* measures the overall incidence of convalescent care programs. The variable has been generated from the following question, which was asked continu-

[7] If the respondent was interviewed in two subsequent waves, e.g., in 1994 and 1995, I match time-variant data from questions posed in the first year dealing with the first year with retrospective data obtained from questions posed in the second year dealing with the first year. For example, in 1995, respondents were asked about their current health status and about their insurance status during the previous year. Hence, I use the 1994 data on health status together with the 1995 data on insurance status if the respondent was interviewed in both years.

ously from 1995 to 1999: “Did you go to a health spa for convalescent care in 199X?” In German, this question is even clearer because of the well-known umbrella term *Kur* and the inpatient treatment this entails at a different location from the recipient’s place of residence, a *Kurort* or spa town, which minimizes measurement errors. The fact that we do not know the exact period of the therapy does not severely hamper the analysis, especially since such treatments are usually not carried out over Christmas or New Year’s. Hence, there should be no doubt as to whether the therapy was in 1996 or in 1997.

While *convalescent care* can be considered a fairly good measure of the incidence of convalescent care treatments, the SOEP does not include a measure of their duration. However, as explained above, the length of treatment is regulated by social law and deviations from it are solely determined by the medical personnel and the SHI sickness fund, not by the patient. Therefore, the empirical analysis focuses mainly on the effects on the incidence, which is the key behavioral parameter in this setting. I use aggregated administrative data on the average duration of treatments as an additional outcome measure in descriptive assessments later on.

In my main empirical models, I make use of various control variables. These control variables capture personal and family-related characteristics such as *age*, *female*, *immigrant*, *partner*, and *children*. Moreover, I control for educational characteristics by using data on the highest school degree obtained. An important determinant of the demand for convalescent care programs is the health status of the respondents, which I observe and control for. I also include covariates that measure whether the person was employed full-time, part-time, marginally, or not at all.[8] I additionally control for gross monthly income and the *equalized household income*, which I obtain by dividing the household income by the root of all household members. To capture time-invariant regional characteristics, I make use of 15 state dummies. Regional labor market dynamics are controlled for by the inclusion of the annual state unemployment rate. Time trends are captured by year dummies. A list of the covariates, their means, and standard deviations can be found in Appendix E.

[8] Non-employment in particular may change quickly. Hence, the assignment of respondents to the treatment and control group might be imprecise. Since the SHI/PHI status is very stable over time, the imprecision lies between the different subgroups that were insured under the SHI as well as between the different subgroups that were insured under the PHI.

5.3.3 Treatment Indicators

In Section 5.2.2, I defined three mutually exclusive subsamples that were affected by different reform elements, as shown in Table 3.1. In the next section, I make use of three distinct models to assess the effectiveness of the various reforms. This requires three distinct treatment indicators for the three models to compare the different subsamples.

T1 has a one for employees in *Treatment Group 1* and a zero for respondents in the *Control Group*. By using this treatment indicator in *Model 1*, I compare those who were affected by all reforms with those who were totally unaffected to assess the net effect of all reforms jointly on the demand for convalescent care programs.

T2 has a one for employees in *Treatment Group 2* and a zero for respondents in the *Control Group*. Thus, in *Model 2*, I contrast those who were affected by the first two reforms with the non-treated. In this model, my main intention is to evaluate the effectiveness of the copayment doubling, i.e., the first reform. In extended robustness checks, I will also assess the effect of the second reform by means of *Model 2*.

T3 is used in *Model 3*, which assesses the effectiveness of the cuts in paid leave. For this purpose I compare *Treatment Group 1* with *Treatment Group 2*. Thereby I extract the effect of the first two reforms from the net reform effect to obtain the effect of the cuts in paid leave.

5.4 Estimation Strategy

5.4.1 Difference-in-Differences

I would like to measure how a certain reform affected the incidence of convalescent care programs. Thinking of the policy intervention as a treatment, I fit a probit model of the form:

$$P[y_{it} = 1|\boldsymbol{X_{it}}] = \Phi(\alpha + \beta post97_t + \gamma T_{it} + \delta \underbrace{(post97_t \times T_{it})}_{DiD_{it}} + \boldsymbol{s'_{it}}\boldsymbol{\psi} + \rho_t + \phi_s + \epsilon_{it}) \quad (5.1)$$

where y stands for the incidence of convalescent care programs, *convalescent care*. *post97* is a dummy that takes on the value one for post-reform years and zero for pre-reform years. Depending on the model, T stands for one the three treatment indicators (see Section 5.3.3 above). The interaction term between the two dummies gives us the difference-in-differences (*DiD*) estimator. To evaluate how the reform affected the outcome variable y, henceforth, I always compute and display the marginal effect of the interaction term $\frac{\Delta\Phi(.)}{\Delta(post97\times T)}$.[9] $\Phi(.)$ is the cumulative distribution function for the standard normal distribution. By including additional time dummies, ρ_t, I control for common time shocks. State dummies, ϕ_s, account for permanent differences across the 16 German states along with the annual state unemployment rate that controls for changes in the tightness of the regional labor market and that is included in the $K \times 1$ column vector $\boldsymbol{s}'_{\boldsymbol{it}}$. The other $K-1$ regressors are made up of personal controls including health status, educational controls, and job-related controls as explained in Section 5.3.1.

5.4.2 Identification

The identification strategy relies on DiD estimation and hence on the assumption of a common time trend of the outcome variable for treatment and control group in the absence of the policy intervention. This assumption should hold conditional on all available covariates. In almost all natural experiments and non-randomized settings, controlling for a rich set of covariates is important since control and treatment group differ with respect to most of the observed characteristics. This is also true in the present case, as Table 5.2 shows.

For example, in comparison to the *Control Group*, *Treatment Group 1* includes more females and immigrants, and the employees are less educated. As compared to the *Control Group*, the people in *Treatment Group 2* are younger and more likely to be full-time employed.

As can be seen from Table 5.3, the most important driver of the demand for convalescent care programs is health status. Not surprisingly, age also plays a role, as well as income. Immigrants are less likely to receive convalescent care, probably because of information asymmetries.

[9] Puhani (2008) has shown that the advice of Ai and Norton (2004) to compute the discrete double difference $\frac{\Delta^2\Phi(.)}{\Delta post97 \Delta T}$ is not of relevance in nonlinear models when the interest lies in the estimation of a treatment effect in a difference-in-differences model. Using treatment indicators, the average treatment effect on the treated is given by $\frac{\Delta\Phi(.)}{\Delta(post97\times T)} = \Phi(\alpha + \beta\text{post97} + \gamma\text{T} + \delta\text{DiD} + \boldsymbol{s'\psi} + \rho + \phi + \epsilon) - \Phi(\alpha + \beta\text{post97} + \gamma\text{T} + \boldsymbol{s'\psi} + \rho + \phi + \epsilon)$ which is exactly what I calculate and present throughout the chapter.

Table 5.2: Variable Means by Treatment and Control Group

Variable	**Treatment Group 1**	**Treatment Group 2**	**Control Group**
Convalescent Care	0.032	0.045	0.032
Personal characteristics			
Female	0.397	0.614	0.364
Age	37	47	45
Age squared	1,693	2,576	2,231
Immigrant	0.220	0.170	0.064
East Germany	0.246	0.309	0.134
Partner	0.798	0.671	0.745
Children	0.470	0.364	0.404
Good health	0.598	0.463	0.609
Bad health	0.106	0.201	0.110
Educational characteristics			
Dropout	0.051	0.073	0.027
Certificate after 8 years' schooling	0.368	0.425	0.206
Certificate after 10 years' schooling	0.319	0.267	0.299
Certificate after 12 years' schooling	0.034	0.025	0.049
Certificate after 13 years' schooling	0.112	0.113	0.387
Certificate degree	0.116	0.077	0.027
Job characteristics			
Full-time employed	0.831	0.197	0.671
Part-time employed	0.130	0.046	0.053
Marginally employed	0.040	0.010	0.008
Civil servant	0.000	0.010	0.427
Public servant	0.000	0.679	0.645
Self employed	0.000	0.056	0.258
Apprentice	0.000	0.057	0.012
Gross income per month	1,860	618	2,126
Regional unemployment rate	11.706	12.317	11.031
N	23,530	4,261	37,758

In contrast to Appendix E, this table gives mean values separately for the treatment and control groups. As detailed in Section 5.3, *convalescent care* is the overall incidence of convalescent care programs.

Table 5.3: Determinants of Convalescent Care

Variable	Coefficient	Standard Error
Personal characteristics		
Female (d)	-0.0006	0.002
Age	0.0023***	0.000
Age squared /1,000	-0.0169***	0.003
Immigrant	-0.0056**	0.002
East Germany	0.0017	0.007
Partner	-0.0025	0.002
Children	-0.0014	0.002
Good health	-0.0230***	0.002
Bad health	0.0398***	0.003
Educational characteristics		
8 years of completed schooling	0.0044	0.004
10 years of completed schooling	0.0100**	0.004
12 years of completed schooling	0.0106	0.007
13 years of completed schooling	0.0046	0.004
Other certificate	0.0045	0.004
Job characteristics		
Full-time employed	0.0017	0.002
Part-time employed	-0.0035	0.003
Marginally employed	-0.0033	0.005
Gross wage per month/1,000	-0.0019**	0.001
Regional unemployment rate	-0.0028***	0.001
R-squared	0.0947	
χ^2	1,542	
N	65,549	

* $p<0.10$, ** $p<0.05$, *** $p<0.01$; marginal effects, which are calculated at the means of the covariates, are displayed. Dependent variable is *convalescent care* and measures the incidence of all convalescent care programs. Standard errors in parentheses are adjusted for clustering on person identifiers. Regression includes state dummies. Left out reference categories are dropout and non-employed.

Again, I would like to stress that the econometric specifications adjust the sample composition to the various personal, educational, and job-related characteristics of the respondents. Recall that the health status of the respondents is observed and controlled for. Likewise, adjustments are made for time effects, persistent differences between states, and the annual state unemployment rate.

The key identifying assumption, the common time trend assumption, is likely to hold. It assumes the absence of unobservables that generate different outcome *dynamics* for the treatment and control group. It is worth mentioning that a selection on observables story is very plausible in the present setting. In the first place, it is the SHI/PHI insurance status that determines treatment (see Table 5.1). Almost all factors that determine whether respondents are insured under the SHI or PHI—such as occupational status and income—are observed.

A method to check the absence of distorting unobservable effects is to estimate placebo regressions for years without a reform. I will make use of this method in the next section.

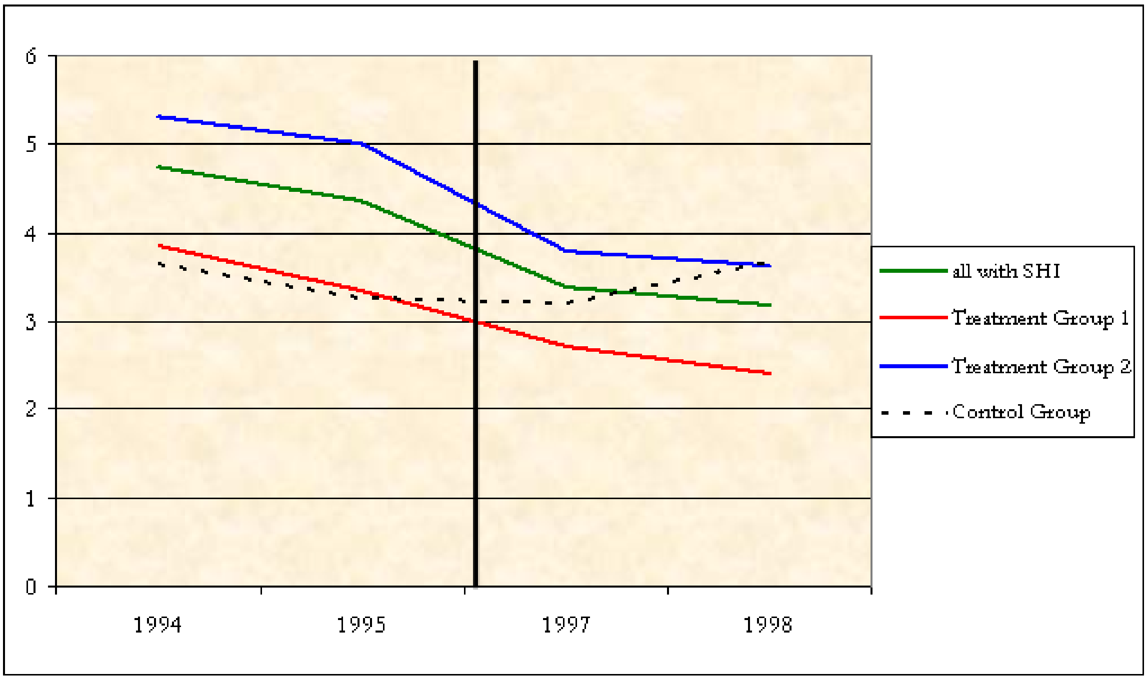
Figure 5.1 shows the evolution of the outcome variable for *Treatment Group 1, 2* and the *Control Group* over time.[10] Even without the correction for observables, we observe a parallel evolution in the three groups during the pre-reform years. After the reform, the incidence of convalescent care programs in the control group remained fairly stable, whereas we observe a clear, distinct, and parallel decrease for the treatment groups.

Compositional changes within the treatment and control groups might have an impact on the outcome variable. For example, in *Treatment Group 2*, the share of self-employed or public-sector employees may change over time, which might affect or even produce the trend in the outcome variable. However, the share of self-employed people within *Treatment Group 2* only fluctuated between 5.31 percent and 5.86 percent from 1994 to 1998. The other subgroups showed similar fluctuations, also remaining very stable over time.

In recent years, the drawbacks and limitations of DiD estimation have been debated extensively. A particular concern is the underestimation of OLS standard errors due to serial correlation in the case of long time horizons as well as unobserved (treatment and control) group effects (Bertrand et al., 2004; Donald and Lang, 2007; Angrist and Pischke, 2009). To cope with the serial correlation issue, we focus on short time horizons. In addition, to provide evidence on whether unobserved common group errors might be a serious threat to our estimates, in robustness checks,

[10] As will be shown later, there is evidence that distorting effects play a role due to the announcement of the reform at the end of 1995. Hence, the two uncontaminated pre-reform years, 1994 and 1995, are contrasted to the two post-reform years, 1997 and 1998.

Figure 5.1: Incidence of Convalescent Care Programs by Year and Subsamples



we cluster on the state × year (16 × 5 = 80 clusters) level (Angrist and Pischke, 2009).

A crucial issue in most studies that try to evaluate policy reforms is, besides the absence of a control group, selection into or out of the policy intervention. I can cope with concerns about selection since I am in the fortunate position of having a framework in which two almost totally independent health care systems exist side by side, as explained in Section 5.2. On the one hand, this provides a well defined control group. On the other hand, I do not need to fear that reform-induced selection has distorted the results, as there is virtually no switching between the SHI and the PHI, and since all SHI-insured persons are covered by universal health plans. Due to strict German regulations, a switch to the PHI was only legally allowed for a small fraction of optionally SHI-insured individuals, and I am able to identify and exclude these cases when running robustness checks. In my dataset, only 1.6 percent of those who were insured under the SHI for at least one year switched to the PHI between 1994 and 1998. The rate did not increase after the reform. Only 1.3 percent of those who were insured under the SHI in 1995 switched to the PHI in 1997 or 1998. We need to consider the possibility of pull-forward effects. Convalescent care programs are usually planned several months or even years in advance. Since the first policy reform plans were made public at the end of 1995 (Handelsblatt, 1995), it may be that a significant portion of the SHI-insured received their convalescent care therapy in 1996 instead of 1997. In the empirical application, I will check for anticipation effects.

Admittedly, it may have been that, due to rising awareness and increased political pressure, the SHI and SPI were more restrictive in their authorization of therapy programs during the period when the reforms were under political discussion, i.e., in 1996. As for anticipation effects that might have been triggered by the insured, one can test for such effects by either excluding the year 1996 from the analysis or by adding an interaction term between 1996 and the treatment indicator to the analysis.

To be able to fully attribute changes in the incidence to changes in the demand for convalescent care programs, supply-side effects should not play a role. I have not found indications of supply-side constraints. In contrast, there have been reports about the deepest crisis in the market for convalescent care since the end of the Second World War (Handelsblatt, 1998). Dozens of medical facilities and health spas have had to close and, hence, there is strong evidence that there was an excess of supply. This is also supported by official statistics stating that the utilized bed capacity of all facilities strongly decreased, from 83.2 percent in 1996 to 62.3 percent in 1997 (German Federal Statistical Office, 2010).

Individuals insured under the SHI who were for some reason exempted from copayments are not identifiable. For example, people whose annual copayments for pharmaceuticals, health care services, or medical devices exceeded a certain percentage of their disposable household income could have applied for a case of hardship.[11] However, at that time, the *German Spa Association* claimed that the public was widely unaware of the exemption clauses. This should therefore not downwardly bias the results severely.

As has already been mentioned in Section 5.2.2, the third reform allowed employers to deduct two days of paid vacation for every five days that an employee was absent from work due to convalescent care therapy. The fourth reform cut statutory (short-term) sick pay. In contrast to the other reforms, these two reforms are rather indirect cost containment measures since they decreased the statutory min-

[11] The usual threshold is 2 percent of disposable household income; for people with chronic diseases it is 1 percent.

imum standards. Since employers are always free to provide fringe benefits on top of statutory requirements, reforms 3 and 4 simply increased employers' capacity to act. I cannot observe which employers enforced these reforms strictly and passed on the decrease in social law minimum standards one-to-one to their employees. Anecdotal evidence and polls suggest that this might have been the case for about 50 percent of all potentially treated, i.e., private-sector employees (Ridinger, 1997; Jahn, 1998). Using all private-sector employees jointly as treatment group, Ziebarth (2009) have shown that the cut in statutory short-term sick pay significantly reduced absenteeism. Since I apply the same approach in this setting, I should be able to identify potential reform effects. Indeed, one of the main objectives of this chapter is precisely to evaluate the effectiveness of direct cost containment measures such as copayment increases, which apply to the entire population, as compared to indirect measures such as decreasing legal minimum requirements, which only increase employers' options to regulate work conditions at the firm level.

As a last point, it should be kept in mind that the identification strategy for the difference-in-differences regression models is based on various specifications. In total, I estimate three distinct models, each of which compares different mutually exclusive and differently affected subsamples. In addition, I run various robustness checks, which enables me to automatically cross-check the consistency and plausibility of the reform effects identified.

5.5 Results

5.5.1 Assessing the reforms' effectiveness

Table 5.4 shows the results for *Model 1, 2*, and *3*. For each model, I display the "raw" difference-in-differences (DiD) estimate as well as the estimates that I obtain from a Probit and an OLS specification with the full set of covariates. The raw estimate represents what we have already seen in Figure 1, which displays the unconditional trends for the various subsamples over time. All models in Table 5.4 use an unbalanced panel, and each column represents one DiD model. *DiD* always stands for the DiD estimate.

Model 1 makes use of the treatment indicator *T1* and compares the pre-post-reform outcome difference for *Treatment Group 1* to the pre-post-reform outcome difference for the *Control Group*. Since *Treatment Group 1* was affected by all four cost containment measures and the *Control Group* by none, *Model 1* estimates the net effect of all reforms on the incidence of convalescent care programs. Column (1) gives the raw estimate, column (2) the Probit estimate, and column (3) the OLS estimate under the inclusion of all covariates.

All three estimates for *Model 1* yield significantly negative reform effects on the incidence of convalescent care programs. Moreover, although the point estimates decrease slightly when the full set of covariates is considered, all three estimates are fairly robust. Especially the Probit and the OLS estimates in columns (2) and (3) are very close to one another. The pre-reform incidence of convalescent care programs for *Treatment Group 1* was 0.0355, i.e., 3.55 percent. Relating the percentage point estimate (-0.0081) from my preferred specification in column (2) to this pre-reform incidence rate suggests that all reforms jointly decreased the demand for convalescent care programs by 22.8 percent.

Model 2 disentangles the effects of reforms 1 and 2 from the effects of reforms 3 and 4. Reform 1 doubled the daily copayments for convalescent care treatments. Reform 2 reduced the legally codified standard length of the therapy and increased the waiting times between two therapies. Reform 3 cut statutory sick leave, and Reform 4 cut paid vacation in case of work absences due to convalescent care treatments. *Model 2* contrasts those who were affected by Reforms 1 and 2 (*Treatment Group 2*) with those who were totally unaffected by all health reforms (*Control Group*). It employs the treatment indicator *T2*.

Again, all three estimates are very similar in magnitude: all are negative and significantly different from zero, they are insensitive to the inclusion of covariates, and the results from the OLS and Probit models barely differ. All *DiD* point estimates are slightly larger than the ones in *Model 1*. The average pre-reform convalescent care incidence for *Treatment Group 2* was 0.0502, and hence the -0.0136 percentage point estimate of the Probit model in column (5) translates into a reform-induced decrease of about -27.1 percent. This suggests that Reforms 1 and 2 were responsible for the decrease in demand for convalescent care programs.

Table 5.4: Assessing the Reforms' Effectiveness: Net Effect, Copayment Effect, and Effect of Cut in Paid Leave

	Model 1: Net effect			**Model 2: Copayment effect**			**Model 3: Cut in paid leave**		
Variable	**Raw**	**Probit**	**OLS**	**Raw**	**Probit**	**OLS**	**Raw**	**Probit**	**OLS**
DiD	**-0.0129****	**-0.0081****	**-0.0096***	**-0.0165*****	**-0.0136****	**-0.0163*****	**0.0035**	**0.0016**	**0.0045**
	(0.0057)	**(0.0041)**	**(0.0057)**	**(0.0057)**	**(0.0056)**	**(0.0057)**	**(0.0031)**	**(0.0024)**	**(0.0031)**
Treatment indicator	0.0046	0.0055***	0.0106**	0.0193***	0.0051*	0.0077*	-0.0147***	0.0010	-0.0058**
(*T1, T2, or T3*)	(0.0042)	(0.0020)	(0.0044)	(0.0042)	(0.0028)	(0.0044)	(0.0023)	(0.0020)	(0.0026)
Post-reform dummy	0.0035	-0.0045	-0.0089	0.0035	-0.0007	-0.0018	-0.0130***	-0.0149***	-0.0214***
(*post97*)	(0.0053)	(0.0041)	(0.0066)	(0.0053)	(0.0050)	(0.0063)	(0.0021)	(0.0029)	(0.0037)
Year 1997 (d)		0.0002	0.0000		0.0012	0.0018		0.0015	0.0015
		(0.0020)	(0.0030)		(0.0021)	(0.0027)		(0.0018)	(0.0021)
Year 1996 (d)		-0.0047**	-0.0083**		-0.0049**	-0.0078**		-0.0051***	-0.0069
		(0.0019)	(0.0038)		(0.0020)	(0.0034)		(0.0017)	(0.0028)
Year 1995 (d)		-0.0021	-0.0036		-0.0022	-0.0035		-0.0020	-0.0028
		(0.0018)	(0.0033)		(0.0019)	(0.0031)		(0.0016)	(0.0025)
Educational covariates	no	yes	yes	no	yes	yes	no	yes	yes
Job covariates	no	yes	yes	no	yes	yes	no	yes	yes
Personal covariates	no	yes	yes	no	yes	yes	no	yes	yes
Regional unempl. rate	no	yes	yes	no	yes	yes	no	yes	yes
State dummies	no	yes	yes	no	yes	yes	no	yes	yes
Year dummies	no	yes	yes	no	yes	yes	no	yes	yes
R-squared	0.0006	0.1217	0.0428	0.0012	0.0901	0.0339	0.0019	0.0956	0.0054
χ^2/F-stat	6	793	12	16	992	19	36	1454.9678	8
N	27,791	27,791	27,791	42,019	42,019	42,019	61,288	61,288	61,288

* p<0.1, ** p<0.05, *** p<0.01; in columns (2), (4), and (6), marginal effects are displayed; they are calculated at the means of the covariates except for *T1 (T2, T3)*(=1) and *DiD*(=1). Dependent variable is *convalescent care* and measures the incidence of convalescent care programs (see Section 5.3). Every column represents one regression model; All columns except for (2), (4), and (6) estimate OLS models. Columns (1) to (3) use *T1*, columns (4) to (6) use *T2*, and columns (7) to (9) use *T3* (see Section 5.3 for further details.)Standard errors in parentheses are adjusted for clustering on person identifiers.

In the robustness checks below, I provide evidence that the copayment doubling was probably responsible for the bulk of this decrease.[12] My findings below suggest that the increase in waiting times did not contribute much to the decrease and that the legally codified reduction in the standard length of treatments merely reduced the average duration of treatments.

Model 3 compares those affected by all four reforms (*Treatment Group 1*) to those affected by Reforms 1 and 2 (*Treatment Group 2*). I thereby assess the effects of Reforms 3 and 4 jointly, i.e., the cuts in paid leave. The results of *Model 3* strongly confirm the findings of *Model 1* and *Model 2*: columns (7) to (9) of Table 5.4 all yield point estimates that are very close to zero and not statistically different from zero. The point estimates are even positive and the standard errors are fairly tight. All in all, I do not find any evidence that the cuts in paid leave induced any significant reduction in the demand for convalescent care programs. I have two explanations for this finding. First, the cut in vacation days may not have been a binding constraint, since many employees use all or part of their paid vacation for convalescent care. Although entitled to take paid leave *in addition* to their paid vacation, many employees fear negative job consequences, especially when unemployment rates are high. Second, the cut in sick pay did not necessarily impose any limitation on the insured since their decision may have been between either going to a convalescent care facility or simply staying home to recover. In any case, the patient would have been on sick leave. If necessary, physicians usually recommend treatments in spa towns, but if patients prefer to stay home on sick leave, their wishes are usually respected.

The entire setup and the fact that all results are based on a comparison of three mutually exclusive subsamples gives rise to another means of calculating the effects for *Model 2* and the first two reforms: one can simply subtract the estimates from *Model 3* from those from *Model 1*, i.e., subtract the effects of Reforms 3 and 4 from the net effect of all reforms. It is easy to see that this exercise yields very consistent alternative estimates for *Model 2* that are almost identical to the direct estimates in columns (5) and (6) (0.0097 for the Probit and 0.0141 for the OLS model).

[12] Ziebarth (2010) has shown that the price elasticity of demand for convalescent care treatments is inelastic and about -0.4.

5.5.2 Robustness checks

Table 5.5 displays various robustness checks. In all cases, I focus on *Model 2* and the Probit specification with all covariates included.[13]

The first column of Panel A is the same estimate as the one in column (5) of Table 5.4 (-0.0136). This is my "standard" estimate. Column (2) excludes the year 1996 from the specification. Since the copayment doubling was first announced in December 1995, is it likely that the pre-reform year 1996 is contaminated by either pull-forward effects triggered by the insured or by supply-side effects triggered by SHI sickness funds or the SPI. The SHI and SPI might have been more restrictive in the authorization of treatments due to rising public awareness and political pressure. Indeed I find some evidence of this. Omitting 1996, the DiD estimate shrinks slightly and translates into a decrease of about 21 percent in demand. Column (3) also supports this result, since the short-run reform effect obtained by comparing 1996 to 1997 is larger than the standard estimate in column (1) and -0.018.

Reform 2 increased the waiting period for SHI-insured from three to four years. The last column in Panel A tests whether the increase in waiting times has reduced the incidence of convalescent care programs in the short run. The waiting period is the time required to elapse between two treatments. However, this extension of the waiting period did not apply to indviduals needing urgent medical treatment. As detailed in Section 5.2, people insured under the SHI have free choice of doctors, and there are almost no waiting times for doctor appointments in Germany. Thus, it is unlikely that this change had a substantial effect, since finding a doctor to write a prescription for treatment is usually not difficult. The increase in the waiting period forced patients who had received treatment in 1994 (1995) to wait until 1998 (1999) instead of 1997 (1998) in the absence of urgent medical reasons. Thus, if the increased waiting period had a substantial impact, I would measure a stronger reform effect for 1997 than for 1998. Column (5) of Table 5.5 shows that the reform effect in 1997 was even lower than in 1998. I take this as evidence that the increased waiting period had no significant (short-term) effect on the demand for convalescent care.

The second element of Reform 2 was the reduction of the standard length of

[13] Here I focus on *Model 2* since it includes more observations than *Model 1* and therefore yields a more precise estimation. Morever, as such, I am able to run checks on the effectiveness of Reform 2. The results for *Model 1* are very similar and available in the working paper version.

convalescent care from four to three weeks. The standard length is codified in the Social Code Book and applies universally to everyone who is insured under the SHI. Exceeding the standard length is only possible in case of urgent medical reasons. The decision to deviate from the legally codified standard length can only be made by the attending physician after consulting the sickness fund to authorize the prolongation.

Since the SOEP does not include information on the length of therapy, I cannot estimate the effect of the reduction in the standard length using a regression model. However, official data is available on the average treatment length and the total number of days spent in inpatient medical facilities for convalescent care treatments. These official data represent average values for the whole of Germany. According to these data, the average treatment length for all insured individuals decreased by almost 4 days from 31.0 (30.2) days in 1995 (1996) to 27.3 (26.4) days in 1997 (1998) (German Federal Statistical Office, 2010). The figures provide descriptive evidence that reducing the legally codified standard length was an effective tool to reduce the real length of treatments. However, it is unlikely that reducing the legal standard length of therapies had a substantial impact on the *incidence* of convalescent care therapy, i.e., on the decision to go to a health spa. However, I have no means to prove this assumption empirically.

Panel B of Table 5.5 presents additional robustness checks. The first three columns prove that treatment selection or panel attrition pose no threat to my results. In the first column, I balance the sample. In column (2), I weight the standard regression with the inverse probability that a respondent did not drop out of my sample in the post-reform period. In the third column, I exclude the only population group from my sample that could have selected themselves out of the treatment. Only respondents who were optionally insured under the SHI system had the possibility by opting out of the SHI. However, as discussed above, opting out is essentially a lifetime decision and therefore very rare. The DiD estimates from all three robustness checks are close in size to the standard estimate in column (1) of Panel A. Each estimate is significantly different from zero.

I exclude health variables in column (4) since the health status might be endogenous if measured after a convalescent care therapy.[14] The resulting estimate is very

[14] Keep in mind that health status refers to the time of the interview, whereas the information about convalescent care programs is sampled retrospectively for the previous calendar year. As explained at the beginning of Section 5.3, if a respondent was interviewed in two subsequent years, I match the current health status information in year t_0 with the convalescent care information from

robust.

Table 5.5: Robustness Checks

Panel A					
	Standard	**w/o 1996**	**'96 vs. '97**	**flexible**	
DiD	-0.0136**	-0.0109*	-0.0180*	DiD98	-0.0115***
	(0.0056)	(0.0062)	(0.0098)		(0.0042)
				DiD97	-0.0089**
					(0.0043)
				DiD96	0.0084
					(0.0080)
Panel B					
	balanced sample	**weighted**	**w/o optionally insured**	**no health covariates**	**cluster state×year**
DiD	-0.0153**	-0.0155**	-0.0125**	-0.0144**	-0.0144**
	(0.0067)	(0.0061)	(0.0057)	(0.0061)	(0.0057)

* p<0.1, ** p<0.05, *** p<0.01; marginal effects are displayed; they are calculated at the means of the covariates except for *T2*(=1) and *DiD*(=1). Dependent variable is *convalescent care* and measures the incidence of convalescent care programs (see Section 5.3). Every cell represents one probit DiD model. All models are similar to the one in column (5) of Table 5.4, i.e., they estimate the copayment effect using Model 2 and comparing *Treatment Group 2* to the *Control Group* (see Section 5.3). Column (1) in Panel A is the standard DiD estimate, i.e., the estimate in column (5) of Table 5.4. Column (2) in Panel A excludes the year 1996 from the regression and is the estimate excluding anticipation effects. Column (3) in Panel A contrasts the year 1996 to the year 1997 and thus estimates the reforms' short-run effect. Column (5) in Panel A shows the most flexibel of all specifications. Instead of interacting the post-reform dummy *post97* with the treatment indicator *T2*, it includes three alternative interaction terms: *Year1996*×*T2* (*DiD96*), *Year1997*×*T2* (*DiD97*), and *Year1998*×*T2* (*DiD98*). Column (1) in Panel B uses a balanced sample and thus excludes panel attrition effects. Column (2) in Panel B weights the regression with the inverse probability that a person does not drop out of the sample in post-reform years. Column (3) in Panel B excludes the only respondents that could have selected themselves out of the treatment, i.e., optionally SHI insured people. Column (4) in Panel B excludes all health measures from the list of covariates. Column (5) in Panel B clusters the standard errors at a higher aggregated level, i.e., the state×year level (80 cluster). Standard errors in all other models are adjusted for clustering on person identifiers and are always in parentheses. All models have 42,019 observations expect for Panel A column (2) (33,975 obs.) and column (3) (16,935) as well as Panel B column (1) (30,625) and column (3) (38,962). For more details about the different model specifications and the interpretation of the results, please see main text.

year t_1 which refers to year t_0. Since two-thirds of all interviews were carried out between January and March, the health status is likely to have been measured before the medical treatment.

The last column in Panel B clusters standard errors on a higher aggregated level to test whether the common group error structure might be a serious issue in this setting (Angrist and Pischke, 2009). As can be seen, there is no evidence of this.

In Table 5.6, I display placebo regressions for *Model 1, 2,* and *3* and Probit as well as OLS specifications. Placebo regressions are a common means to test the common time trend assumption. Finding significant reform effects for years without a reform would cast serious doubts on the plausibility of the common time trend assumption. I use 1994 and 1995 as pseudo-reform years and, apart from that, the same setup as above. All twelve placebo regression estimates are close to, and not significantly different from, zero.

5.5.3 Reduction in health expenditures

Since reducing health expenditures was the main intention behind the policy reforms, I perform a rough calculation of the decrease in public health expenditures using official data. Official data is available on the total sum that was spent on convalescent care by the public social insurance. Taking the simple difference in expenditures in 1997/1998 vs. 1994/1995 yields a total savings estimate of €835 million per year. This represents a decrease in spending of -12.5 percent (German Federal Statistical Office, 2010). It should be kept in mind, however, that time trends are included in this rough savings estimate.

Since copayments were doubled, this reform has raised additional revenues. However, official data show that the total number of convalescent care days consumed decreased by 22 percent from 57 million in 1994/1995 to 44.5 million in 1997/1998 (German Federal Statistical Office, 2010). Multiplying each sum by the pre- and post-reform copayments and taking the difference suggests that increasing copayments not only dampened the demand for convalescent care very effectively but also raised additional revenues of about €435 million per year.[15]

[15] Under the assumption that 18.8 percent of all therapies were undertaken by East Germans (German Federal Statistical Office, 2010) who were charged lower copayments (see Section 5.2.2 for details.)

Table 5.6: Placebo Reform Estimates

	Model 1: Net effect		**Model 2: Copayment effect**		**Model 3: Cut in paid leave**	
Variable	**Probit**	**OLS**	**Probit**	**OLS**	**Probit**	**OLS**
DiD95	0.0015	0.0004	0.0052	0.0058	-0.0029	-0.0058
	(0.0055)	(0.0068)	(0.0076)	(0.0068)	(0.0026)	(0.0039)
DiD94	-0.0004	-0.0008	0.0013	0.0045	-0.0004	-0.0037
	(0.0050)	(0.0073)	(0.0069)	(0.0071)	(0.0026)	(0.0040)
Educational covariates	yes	yes	yes	yes	yes	yes
Job covariates	yes	yes	yes	yes	yes	yes
Personal covariates	yes	yes	yes	yes	yes	yes
Regional unempl. rate	yes	yes	yes	yes	yes	yes
State dummies	yes	yes	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes	yes	yes

* p<0.1, ** p<0.05, *** p<0.01; in columns (1), (3), and (5), marginal effects are displayed; they are calculated at the means of the covariates except for *T1 (T2, T3)*(=1) and *DiD94* (*DiD95*) (=1). All columns but (2), (4), and (6) estimate OLS models. The dependent variable is *convalescent care* and measures the incidence of convalescent care programs (see Section 5.3). Every cell represents one regression model. Columns (1) and (2) use *T1*, columns (3) and (4) use *T2*, and columns (4) and (5) use *T3* (see Section 5.3 for further details). Each model in columns (1) and (2) has 27,791 observations; each model in columns (3) and (4) has 42,019 observations and columns (5) and (6) are based upon 61,288 observations. All models compare the same groups of (pseudo) treated and (pseudo) non-treated respondents than the non-placebo models. *DiD94* (*DiD95*) is an interaction term between the treatment indicator (*T1*, *T2*, or *T3*) and the year 1994 (1995). Standard errors in parentheses are adjusted for clustering on person identifiers.

5.6 Discussion and Conclusion

In this chapter, I have empirically assessed the effectiveness of different cost containment measures within a unifying framework. In Germany, from 1997 on, four different health reforms were implemented to dampen the demand for convalescent care therapies, to fight moral hazard, and to decrease public health expenditures. At that time, experts claimed that around a quarter of all convalescent care therapies were unnecessary (Schmitz, 1996; Sauga, 1996). In 1995, the German public social insurance system spent €7.6 billion for 1.9 million convalescent care treatments.

Two of the health care cost containment measures applied solely, but universally, to those insured with public health insurance. In Germany, public health insurance coexists with private health insurance providers. Strict legal regulations prevent switching between the two independent systems. Privately insured people were not affected by the two reforms and I can address concerns about treatment selection. Moreover, since the other two of the four cost containment measures only applied to employees in the private sector, I am able to define various subsamples that were affected differently by the reforms. Hence, my empirical findings are based on various difference-in-differences models that compare different mutually exclusive subsamples.

The consistency of the findings from these models, together with several robustness checks, allow me to conclude the following: first, all reforms together decreased the demand for convalescent care therapies by about 20 percent. Second, doubling the daily copayments for convalescent care treatments was by far the most effective cost containment measure. This measure was responsible for the major part of the total decline in demand. Third, descriptive evidence from official data suggests that a legally codified reduction in the standard length of the therapies was effective in reducing the true length of the therapies. Fourth, I do not find evidence that increasing the waiting times between two treatments had any significant effect on the *decision* to go for convalescent care. Fifth, while all these policy measures applied universally to every publicly insured person, two other measures evaluated here applied in a rather indirect way. They reduced statutory minimum standards and increased the employers' options to set firm-specific employment conditions. The first of these indirect measures allowed employers to deduct two days of paid vacation for every five days that an employee was unable to work due to a convalescent care therapy. The second cut statutory sick pay for which employees are eligible during convalescent care treatments. I do not find any evidence that these soft cost containment measures were effective in reducing the demand for convalescent care programs. These findings allow me to conclude that, sixth, indirect measures that reduce statutory minimum conditions in the labor market are far less effective in achieving a specific predetermined policy goal; direct measures that apply universally are much more effective.

As a last exercise, using administrative data, I roughly calculate the reduction in public health expenditures that was induced by all reforms. My back-of-the-envelope calculations suggest that public health expenditures decreased by about € 800 million (-12.5 percent) per year due to the decline in the demand for convalescent care. Moreover, doubling the daily copayments raised additional revenues of about € 400 million per year.

The question to what degree such policy reforms succeed in reducing moral hazard or whether they actually lead to adverse health outcomes is difficult to quantify and is beyond the scope of this chapter. The overall decrease in demand fits well with the a priori claims by health experts that a quarter of all pre-reform therapies were unnecessary. Although it is unlikely that moral hazard was totally eliminated by the reforms, it is probable that the majority of the decrease is due to a reduction in moral hazard and led to greater efficiency in the convalescent care market. On the other hand, if medically necessary therapies were not provided, this may have led to adverse health outcomes and, in the long run, to even higher overall health expenditures. Especially in the case of convalescent care, it is difficult to balance the prevailing degree of moral hazard against potential long-term health improvements that may reduce health expenditures and exert positive external effects. Some studies have found positive health effects of health spa stays: patients with chronic diseases experienced reductions in pain and blood pressure, and for a sample of employees, beneficial effects on physical and particularly mental health, such as improved sleep quality, were found (Sekine et al., 2006; Cimbiz et al., 2005; Constant et al., 1998). While two of these studies are purely correlation-based, Constant et al. (1998) estimate the short-term effects of a randomized trial on 224 patients with chronic lower back pain. However, I am not aware of studies that evaluate the long-term health effects of convalescent care therapies. Assessing the long-term effects of health care on health outcomes as well as on health expenditures is a promising field for future research.

Appendix E

Table 5.7: Descriptive Statistics for the Working Sample

Variable	Mean	Std. Dev.	Min.	Max.	N
Dependent variable					
Convalescent care	0.0393	0.1943	0	1	65,549
Covariates					
Treatment Indicators					
T1	0.8467	0.3603	0	1	27,791
T2	0.8986	0.3019	0	1	42,019
T3	0.3839	0.4863	0	1	61,288
Personal characteristics					
Female	0.5195	0.4996	0	1	65,549
Age	44.2602	16.6593	18	99	65,549
Age squared	2236	1625	324	9801	65,549
Immigrant	0.1811	0.3851	0	1	65,549
East Germany	0.2751	0.4465	0	1	65,549
Partner	0.7214	0.4483	0	1	65,549
Children	0.4045	0.4908	0	1	65,549
Good health	0.521	0.4996	0	1	65,549
Bad health	0.1607	0.3672	0	1	65,549
Educational characteristics					
Drop out	0.062	0.2412	0	1	65,549
8 years of completed schooling	0.3905	0.4879	0	1	65,549
10 years of completed schooling	0.2878	0.4528	0	1	65,549
12 years of completed schooling	0.0298	0.1701	0	1	65,549
13 years of completed schooling	0.1305	0.3369	0	1	65,549
Other certificate	0.0878	0.2829	0	1	65,549
Job characteristics					
Full-time employed	0.455	0.498	0	1	65,549
Part-time employed	0.0767	0.2662	0	1	65,549
Marginally employed	0.0204	0.1414	0	1	65,549
Gross wage per month	1,162	1,301	0	51,129	65,549
Regional unemployment rate	12.0	3.9	7	21.7	65,549

Bibliography

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

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

Ai, C. and E. C. Norton (2004). Interaction terms in logit and probit models. *Economics Letters 80*(1), 123–129.

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

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

Badura, B., H. Schröder, and C. Vetter (2008). *Fehlzeiten-Report 2007: Arbeit, Geschlecht und Gesundheit* (1 ed.). Springer Medizin Verlag.

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

Bazzoli, G. J., H.-F. Chen, M. Zhao, and R. C. Lindrooth (2008). Hospital financial condition and the quality of patient care. *Health Economics 17*(8), 977–995.

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.

Bishai, D., J. Sindelar, E. Ricketts, S. Huettner, L. Cornelius, J. Lloyd, J. Havens, C. Latkin, and S. Strathdee (2008). Willingness to pay for drug rehabilitation: implications for cost recovery. *Journal of Health Economics 27*(4), 959–972.

Bonato, L. and L. Lusinyan (2004). Work absence in Europe. IMF Working Paper 04/193, IMF. http://imf.org/external/pubs/ft/wp/2004/wp04193.pdf, last accessed at December 19, 2008.

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

Brors, P. and P. Thelen (1998). Neue Runde im Streit um die Lohnfortzahlung. *Handelsblatt 59: 25.03.1998*, 3.

Brown, S. (1994). Dynamic implications of absence behaviour. *Applied Economics 26*, 1163–1175.

Brown, S. and J. G. Sessions (1996). The economics of absence: theory and evidence. *Journal of Economic Surveys 10*(1), 23–53.

Burkhauser, R. V., J. S. Butler, and G. Gumus (2004). Dynamic programming model estimates of social security disability insurance application timing. *Journal of Applied Econometrics 19*(6), 671–685.

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

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

Card, D. and A. B. Krueger (1994). Wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review 84*(4), 772–793.

Chandra, A. and A. A. Samwick (2005). Disability risk and the value of disability insurance. NBER Working Papers, National Bureau of Economic Research, Inc.

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.

Chiappori, P.-A., F. Durand, and P.-Y. Geoffard (1998). Moral hazard and the demand for physician services: first lessons from a natural experiment. *European Economic Review 42*(3-5), 499–511.

Cimbiz, A., V. Bayazit, H. Hallaceli, and U. Cavlak (2005). The effect of combined therapy (spa and physical therapy) on pain in various chronic diseases. *Complementary Therapies in Medicine 13*(4), 244–250.

Civan, A. and B. Koksal (2010). The effect of newer drugs on health spending: do they really increase the costs? *Health Economics 19*(5), 581–595.

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

Cockx, B. and C. Brasseur (2003). The demand for physician services: evidence from a natural experiment. *Journal of Health Economics 22*(6), 881–913.

Constant, F., F. Guillemin, J. F. Collin, and M. Boulangé (1998). Use of spa therapy to improve the quality of life of chronic low back pain patients. *Medical Care 36*(9), 1309–1314.

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.

de Jong, P., M. Lindeboom, and B. van der Klaauw (2010). Screening disability insurance applications. *Journal of the European Economic Association*. forthcoming.

Deb, P. (2001). A discrete random effects probit model with application to the demand for preventive care. *Health Economics 10*(5), 371–383.

Deb, P. and P. K. Trivedi (1997). Demand for medical care by the elderly: a finite mixture approach. *The Journal of Applied Econometrics 12*(3), 313–336.

Deb, P. and P. K. Trivedi (2009). Provider networks and primary-care signups: do they restrict the use of medical services? *Health Economics 18*(12), 1361–1380.

Di Matteo, L. (2005). The macro determinants of health expenditure in the United States and Canada: assessing the impact of income, age distribution and time. *Health Policy 71*(1), 23–42.

Dionne, G. and B. Dostie (2007). New evidence on the determinants of absenteeism using linked employer-employee data. *Industrial & Labor Relations Review 61*(1), 108–120.

Doherty, N. (1979). National insurance and absence from work. *The Economic Journal 89*(353), 50–65.

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.

Dusheiko, M., H. Gravelle, R. Jacobs, and P. Smith (2006). The effect of financial incentives on gatekeeping doctors: evidence from a natural experiment. *Journal of Health Economics 25*(3), 449–478.

Ellis, R. P. and T. G. McGuire (1996). Hospital response to prospective payment: moral hazard, selection, and practice-style effects. *Journal of Health Economics 15*(3), 257–277.

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.

Felder, S. (2008). To wait or to pay for medical treatment? Restraining ex-post moral hazard in health insurance. *Journal of Health Economics 27*(6), 1418–1422.

Fenn, P. (1981). Sickness duration, residual disability, and income replacement: an empirical analysis. *The Economic Journal 91*(361), 158–173.

Frick, J. R. and M. M. Grabka (2005). Item-non-response on income questions in panel surveys: incidence, imputation and the impact on inequality and mobility. *Allgemeines Statistisches Archiv 89*(1), 49–60.

Gerfin, M. and M. Schellhorn (2006). Nonparametric bounds on the effect of deductibles in health care insurance on doctor visits – Swiss evidence. *Health Economics 15*(9), 1011–1020.

German Federal Statistical Office (1995). *Statistical Yearbook 1995 for the Federal Republic of Germany.* Metzler-Poeschel.

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

German Federal Statistical Office (1997). *Statistical Yearbook 1997 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 (2000). *Statistical Yearbook 2000 for the Federal Republic of Germany.* Metzler-Poeschel.

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

German Federal Statistical Office (2008). *Finanzen und Steuern: Personal des öffentlichen Dienstes 2007.* Fachserie 14, Reihe 6.

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

German Federal Statistical Office (2009b). *Statistical Yearbook 2009 For the Federal Republic of Germany.* Metzler-Poeschel.

German Federal Statistical Office (2010). *Federal Health Monitoring.* www.gbe-bund.de, last accessed at June 25, 2010.

German Ministry of Health (2008). www.bmg.bund.de, last accessed at 22.02.2008.

German Ministry of Health (2009). Gesetzliche Krankenversicherung: Krankenstand 1970 bis 2008. www.bmg.bund.de, last accessed at 28.01.2010.

German Statutory Pension Insurance (2008). *The German Statutory Pension Insurance as time series.* http://forschung.deutsche-rentenversicherung.de, last accessed at August 14, 2009.

Goldman, D. P., S. D. Hosek, L. S. Dixon, and E. M. Sloss (1995). The effects of benefit design and managed care on health care costs. *Journal of Health Economics 14*(4), 401–418.

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

Handelsblatt (1995). Reformpaket der Koalition. Kur soll künftig Urlaub kosten. *Handelsblatt 19.12.1995*, 3.

Handelsblatt (1998). Gesundheitsreform sorgt für Aderlaß. *Handelsblatt 10.03.1998*, 21.

Hans Böckler Stiftung (2009). *WSI Tarifarchiv.* www.boeckler.de, last accessed at November 12, 2009.

Hans Böckler Stiftung (2010). *WSI Tarifarchiv.* www.boeckler.de, last accessed at August 12, 2010.

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

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

Herring, B. (2010). Suboptimal provision of preventive healthcare due to expected enrollee turnover among private insurers. *Health Economics 19*(4), 438–448.

Hill, S. C. and B. L. Wolfe (1997). Testing the HMO competitive strategy: an analysis of its impact on medical care resources. *Journal of Health Economics 16*(3), 261–286.

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

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

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

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

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

Johansson, P. and M. Palme (2005). Moral hazard and sickness insurance. *Journal of Public Economics 89*(9-10), 1879–1890.

Karlsson, M. (2007). Quality incentives for GPs in a regulated market. *Journal of Health Economics 26*(4), 699 – 720.

Keeler, E. B., G. Carter, and J. P. Newhouse (1998). A model of the impact of reimbursement schemes on health plan choice. *Journal of Health Economics 17*(3), 297–320.

Keeler, E. B., W. G. Manning, and K. B. Wells (1988). The demand for episodes of mental health services. *Journal of Health Economics 7*(4), 369–392.

Keeler, E. B. and J. E. Rolph (1988). The demand for episodes of treatment in the Health Insurance Experiment. *Journal of Health Economics 7*(4), 337–367.

Kenkel, D. S. (2000). Prevention. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics*, Volume 1 of *Handbook of Health Economics*, Chapter 31, pp. 1675–1720. Elsevier.

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

Lambsdorff, O. G. (1982). Konzept für eine Politik zur Überwindung der Wachstumsschwäche und zur Bekämpfung der Arbeitslosigkeit. Dokumentation, Neue Bonner Depeche 9/82. http://www.archive.org/details/Lambsdorff-Papier, last accessed at March 20, 2009.

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.

Lee, C. (1995). Optimal medical treatment under asymmetric information. *Journal of Health Economics 14*(4), 419–441.

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.

Mariñoso, B. G. and I. Jelovac (2003). GPs' payment contracts and their referral practice. *Journal of Health Economics 22*(4), 617–635.

Medizinischer Dienst der Krankenversicherung (MDK) (2008). www.mdk.de, last accessed at October 23, 2008.

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

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

Meyerhoefer, C. D. and S. H. Zuvekas (2010). New estimates of the demand for physical and mental health treatment. *Health Economics 19*(3), 297–315.

Müller, R., D. Hebel, B. Braun, R. Beck, U. Helmert, G. Marstedt, and H. Müller (1998). *Auswirkungen von Krankengeld-Kürzungen: Materielle Bestrafung und soziale Diskriminierung chronisch erkrankter Erwerbstätiger. Ergebnisse einer Befragung von GKV-Mitgliedern* (2 ed.). Schriftenreihe zur Gesundheitsanalyse, Volume 1. GEK Edition.

Newhouse, J. P. (1992). Medical care costs: how much welfare loss? *Journal of Economic Perspectives 6*(3), 3–21.

OECD (2006). *OCED Health Data 2006*.

O'Grady, K. F., W. G. Manning, J. P. Newhouse, and B. R. H. (1985). The impact of cost sharing on emergency department use. *The New England Journal of Medicine 313*(8), 484–490.

Okunade, A. A. (2004). Concepts, measures, and models of technology and technical progress in medical care and health economics. *The Quarterly Review of Economics and Finance 44*(3), 363–368.

Pauly, M. V. and F. E. Blavin (2008). Moral hazard in insurance, value-based cost sharing, and the benefits of blissful ignorance. *Journal of Health Economics 27*(6), 1407–1417.

Pettersson-Lidbom, P. and P. Skogman Thoursie (2008). Temporary disability insurance and labor supply: evidence from a natural experiment. Working paper, Stockholm University, Department of Economics. http://people.su.se/ pepet/tdi.pdf, last accessed at March 19, 2008.

Propper, C., S. Burgess, and K. Green (2004). Does competition between hospitals improve the quality of care? hospital death rates and the NHS internal market. *Journal of Public Economics 88*(7-8), 1247–1272.

Propper, C., B. Croxson, and A. Shearer (2002). Waiting times for hospital admissions: the impact of GP fundholding. *Journal of Health Economics 21*(2), 227–252.

Puhani, P. A. (2008). The treatment effect, the cross difference, and the interaction term in nonlinear "difference-in-differences" models. IZA Discussion Paper Series 3478, IZA. http://www.iza.org, last accessed at 22.02.2008.

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.

Rauch, A., J. Dornette, M. Schubert, and J. Behrens (2008). Berufliche Rehabilitation in Zeiten des SGB II. *IBA-Kurzbericht 25*.

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

Ringel, J. S., S. D. Hosek, B. A. Vollaard, and S. Mahnovski (2002). *The Elasticity of Demand for Health Care: A Review of the Literature and Its Application to the Military Health System* (1 ed.). Rand Corp.

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

Roddy, P. C., J. Wallen, and S. M. Meyers (1986). Cost sharing and use of health services: the United Mine Workers' of America health plan. *Medical Care 24*(9), 873–877.

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

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

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

Sauga, M. (1996). Gesundheit im eigenen Bett: Bonn will den Anstieg der Kur-Ausgaben begrenzen. Die Bäderlobby macht mobil. *Handelsblatt 11.01.1996*, 21.

Schmitz, H. (1996). Politiker, Wissenschaftler und Praktiker suchen Wege, die Verschwendung bei Rehabilitation und Heilmaßnahmen zu stoppen. Kuren sollten für Arbeitnehmer kein preiswerter Urlaub sein. *Handelsblatt 01.04.1996*, 2.

Schreyögg, J. and M. M. Grabka (2010). Copayments for ambulatory care in Germany: a natural experiment using a difference-in-difference approach. *European Journal of Health Economics 11*(3), 331–341.

Schut, F. T. and W. P. M. M. V. de Ven (2005). Rationing and competition in the Dutch health-care system. *Health Economics 14*(S1), S59–S74.

Sekine, M., A. Nasermoaddeli, H. Wang, H. Kanayama, and S. Kagamimori (2006). Spa resort use and health-related quality of life, sleep, sickness absence and hospital admission: the Japanese civil servants study. *Complementary Therapies in Medicine 14*, 133–143.

Shapiro, C. and J. E. Stiglitz (1974). Equilibrium unemployment as a worker discipline device. *American Economic Review 74*(3), 433–444.

Siciliani, L., A. Stanciole, and R. Jacobs (2009). Do waiting times reduce hospital costs? *Journal of Health Economics 28*(4), 771–780.

Sloan, F. A., G. A. Picone, D. H. TaylorJr., and S.-Y. Chou (2001). Hospital ownership and cost and quality of care: is there a dime's worth of difference? *Journal of Health Economics 20*(1), 1–21.

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

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

van Kleef, R., W. van de Ven, and R. van Vliet (2009). Shifted deductibles for high risks: more effective in reducing moral hazard than traditional deductibles. *Journal of Health Economics 28*(1), 198–209.

Voorde, C. V. D., E. V. Doorslaer, and E. Schokkaert (2001). Effects of cost sharing on physician utilization under favourable conditions for supplier-induced demand. *Health Economics 10*(5), 457–471.

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.

Wedig, G. J. (1988). Health status and the demand for health: results on price elasticities. *Journal of Health Economics 7*, 151–163.

Winkelmann, R. (2004). Health care reform and the number of doctor visits – an econometric analysis. *Journal of Applied Econometrics 19*(4), 455–472.

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

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review 93*(2), 133–138.

Wooldridge, J. M. (2006). Cluster-sample methods in applied econometrics: an extended analysis. Working paper, Michigan State University, Department of Economics. https://www.msu.edu/ ec/faculty/wooldridge/current research/clus1aea.pdf, last accessed at March 19, 2009.

Wooldridge, J. M. (2007). What's new in econometrics? Imbens/Wooldridge lecture notes; Summer Institute 2007, lecture 10: Difference-in-differences estimation, NBER. http://www.nber.org/minicourse3.html, last accessed at March 19, 2009.

Ziebarth, N. R. (2009). Long-term absenteeism and moral hazard – evidence from a natural experiment. DIW Discussion Papers 888, German Institute for Economic Research (DIW). http://www.diw.de/documents/publikationen/73/97949/dp888.pdf, last accessed at January 13, 2010.

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

Ziebarth, N. R. and M. Karlsson (2009). A natural experiment on sick pay cuts, sickness absence, and labor costs. SOEPpapers 244, German Institute for Economic Research (DIW). http://www.diw.de, last accessed at December 15, 2009.

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

Zweifel, P. and W. G. Manning (2000). Moral hazard and consumer incentives in health care. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics* (1 ed.), Volume 1, Chapter 8, pp. 409–459. Elsevier.