IFAU
Institute for Evaluation of Labour
Market and Education Policy

Essays on health shocks and social insurance

Martin Nilsson

DISSERTATION SERIES 2014:4
Presented at the Department of Economics, Uppsala University

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P O Box 513, 751 20 Uppsala
Visiting address: Kyrkogårdsgatan 6, Uppsala
Phone: +46 18 471 70 70
Fax: +46 18 471 70 71
ifau@ifau.uu.se
www.ifau.se

This doctoral dissertation was defended for the degree of Doctor in Philosophy at the Department of Economics, Uppsala University, August 29, 2014. Essay 1 is a revised version of IFAU Working paper 2011:11 and Essay 3 has previously been published by IFAU as Working paper 2012:3.

ISSN 1651-4149

Acknowledgements

I owe my gratitude to many people, however, there are a few who have been especially important during the work on this thesis.

First of all, I am thankful to my advisors, Per Johansson and Johan Vikström. Your enthusiasm, suggestions and great knowledge have undoubtedly improved the quality of this thesis. I have often walked to a meeting with Per or Johan full of anguish and self-doubt, only to emerge after an hour fild with new hope and with renewed confidence. Needless to say, your guidance and support during the writing of this thesis have been invaluable.

Furthermore, I want to thank Matti Sarvimäki for carefully reading my papers and give insightful suggestions at my final seminar. For always providing quick and excellent administrative support, I thank the entire administration of both the Department of Economics and IFAU. I also want to thank the Jan Wallander and Tom Hedelius foundation for the financial support that allowed me to spend a semester at University of Maryland and the Maryland Population Research Center (MPRC). I would like to thank Prof. John Ham for inviting me to University of Maryland.

During my time as a graduate student I have made many friends both at the department and at IFAU, and I would like to thank you all. Without you, my time as a graduate student would not have been nearly as fun. A special thank you goes to Mattias, Arizo and Anna.

I also owe great gratitude to my friends and family. Mum and Dad, Jonas, your love and unconditional support have been invaluable and mean more than you probably know.

Finally, all the love goes to Clare who never stopped believing in me. Thank you for making it all worthwhile.

Washington D.C., May 2014
Martin Nilsson

Contents

Introduction

This thesis consists of four self-contained essays in empirical economics. The first paper in this thesis, **Essay 1**, addresses the relationship between health and socioeconomic status. The purpose of the paper is to investigate whether the, so called, socioeconomic gradient in health arises from socioeconomic heterogeneity in the impact of health shocks on labor market outcomes. The second and third paper relates, in different ways, to the organization of the public sickness and disability insurance system. **Essay 2** analyzes a reform of the compulsory Swedish sickness insurance. Specifically, I analyze the indirect impact of replacement rates on medium- and long-term sickness absence. **Essay 3** examines the effect of poor quality medical certificate on sickness absence durations. Finally, the **Essay 4** investigates how socioeconomic status is associated with road accidents. This study also calculates the societal costs of road accidents in terms of lost production.

Heterogeneity in the Impact of Health Shocks on Labor Outcomes

Essay 1 addresses one of the most widely replicated observations in the social sciences, the socioeconomic gradient in health.[1] Within countries, the evidence shows that in general the lower an individual's socioeconomic position the worse their health. In modern societies, the gradient is usually found to widen during working life but then narrows as people reach older ages (see e.g. van Kippersluis et al., 2009; Case & Deaton 2005b).

One suggested explanation to the observed health gradient is the hypothesis that health outcomes influence socioeconomic status. Surveys by Smith (1999) and Case & Deaton (2005a) argue that a larger part of the association between health and socioeconomic status at middle and older ages likely reflects an impact of health on socioeconomic status. The conclusion of little or no effect of income on health outcomes is supported by Adams et al., (2003) and Frijters et al., (2005). Most of the recent literature has implicitly assumed that the impact of a health event on labor market outcomes is similar across different socioeconomic groups. Two exceptions are Smith (1999) and García-Gómez et al., (2013), who both find heterogeneity in the impact of health shocks by income.

[1] See Antonovsky (1967) for a review of the early literature on the socioeconomic gradient in health.

If substantial heterogeneity exists in the effect of health shocks, this could be an important part of the explanation of how the socioeconomic gradient in health arises. This paper advances the recent literature by focusing on heterogeneity by education in the relative effect of health shocks on labor outcomes. In addition, we will allow the effects to vary by age.

To provide new evidence on heterogeneity in the impact of health shocks, this paper uses register-based data on the entire population of Swedish workers. We formulate a Difference-in-Difference design, where we compare the change in labor earnings across matched high and low-educated workers who experience the same type of health shocks. This design takes away the requirement to define control groups of workers not suffering from health shocks.

Our results suggest large heterogeneity in the effects, where low-educated individuals suffer relatively more from a given health shock. These results hold across different types of health shocks and become more pronounced by age. This suggests that at older ages, there are very large differences in the long -run possibilities to cope with a negative health shock. These results are consistent with the idea that the socioeconomic gradient in health is partly caused by the impact of health shocks on socioeconomic status. They also offer one explanation for why the socioeconomic gradient in health widens during middle ages. Furthermore, low-educated workers are also more likely to leave the labor force and take up disability insurance, sickness insurance, and unemployment benefits following a health shock. Our results suggest that socioeconomic heterogeneity in the effect of health shocks offers one explanation to how the socioeconomic gradient in health arises

Economic incentives and sickness absence behavior

Essay 2 analyzes a reform of the compulsory Swedish sickness insurance. Specifically, I empirically analyze the indirect impact of replacement rates on long-term sickness absence.

The presence of moral hazards presents policy makers with a delicate problem: how do you balance the advantageous income-distribution properties of a generous sickness insurance (SI) system with the disadvantageous behavioral responses to such a system?

In recent years, reductions in SI replacement levels have been widely used as a way of lowering absence rates. The cuts are justified by the assumption that economic incentives affect an individual's sick-leave decision. Increasing the cost of being absent would thus lower the absence rate. There is also considerable evidence showing that economic incentives influence take up rates of SI and the duration of short-term absence (see e.g. Meyer et al., 1995; Curington, 1994; Johansson & Palme, 1996, 2002; 2005; Henrekson & Persson, 2004; Ziebarth & Karlsson, 2013; Petterson-Lidbom & Thoursie, 2013). However, less is known about the impact of replacement rates on long-term absenteeism.[2]

To provide new knowledge on this topic, I utilize a reform of the compulsory Swedish SI scheme. Replacement levels were reduced, for all new absence spells, from 90 percent of foregone earnings to 65 percent during the 3 first days in an absence spell and to 80 percent for day 4 to 90 for all new absence spells. From day 91 onwards the rate remained at 90 percent. The reason behind maintaining the compensation level for spells of more than 90 days, was to avoid negative economic effects on an already disadvantaged group.

I show that the reform had an indirect effect on absence behavior. The indirect effect (i.e. the cost of returning to work) increased due to the risk of having to start a new absence spell and receive the lower replacement rate. Using detailed data on the complete account of all sickness absence spells during the period, I am able to identify and estimate the indirect effect on absence-to-work transitions by comparing absence behavior before and after the reform. The empirical findings confirm previous research about economic incentives and absence behavior; I find that the indirect effect of the reform reduced absence-to-work transitions both among medium as well as long-term absentees. In comparison with previous evaluations of the reform, the effect on long-term absence is in line with Johansson & Palme (2005). However, unlike Johansson & Palme I am also able to show that the indirect effect of the reform on medium absence spells is substantial and even greater than for the longer spells.

I also find that the indirect effect on absence-to-work transitions is stronger among white collar workers than among blue collar workers. The difference is not surprising given that white collar workers were covered by supplementary sickness compensation through their collective bargaining agreement before the reform but not after. This meant that the cost of returning to work was amplified for white collar workers. Furthermore, I find that male claimants react more strongly to the reform than women. This is especially true for medium-term absenteeism. This result supports previous findings that men, on average, tend to react more strongly to economic incentives in an SI scheme (see e.g. Henrekson & Persson, 2004; Johansson & Palme, 1996 and Ziebarth & Karlsson, 2013).

[2]Two exceptions to this are Ziebarth (2013) and Hesselius & Persson (2007). In Ziebarth (2013), the author shows that a reduction of the replacement rate in Germany had no effect on long-term absenteeism. For Sweden, Hesselius & Persson (2007) show that an increase in the replacement rate for absence spells lasting longer than 90 days increased the duration of spells.

Should Sickness Insurance and Health Care be Administered by the Same Jurisdiction?

Essay 3 examines the effect of poor quality medical certificate on sickness absence durations.

Sweden has obligatory sickness and disability insurance which is both financed and administered by the government. A person is entitled to sickness benefits only if she or he has a disease or injury that results in reduced capacity to work. In Sweden, the first seven days of a sick-leave spell can be self-certified. After that, insured individuals must have certificates issued by a medical doctor in order to receive sickness benefits. The decisions regarding entitlement to sickness benefits and additional rehabilitation measures are made by the social insurance officers, largely based on the information provided on medical certificates. The medical certificate thus has substantial impact on the life and work situation of the claimants, as well as on the economic costs to the society. Since health care is administered at the county level, this means that monitoring is, to some extent, decentralized at a lower jurisdictional level than the funding and governance of the insurance. This paper studies one consequence of this decentralization: the effect on individual sickness absence when such certificates are not approved by the Swedish Social Insurance Agency (SIA) and are instead re-remitted to the doctor for completion and, potential, approval by the SIA. Previous research has found that physicians frequently found it problematic to handle sickness certification tasks and that the certificates issued often is insufficient and of low quality (see e.g. Wahlström & Alexanderson, 2004; Hussey et al., 2004; Söderberg & Alexanderson, 2005; Löfgren et al., 2007). However, as far as we know, this is the first paper that provides some evidence about the consequences of medical certificate quality on sickness absence.

We find that this re-remission increases the length of sickness absence spells by an average of 30 percent. Our findings do not, however, imply that SIA's role as gatekeepers in the social insurance system should be removed. Without this control mechanism the moral hazard problem associated with the insurance scheme would increase dramatically. This would have a long run negative effect on both the incidence and the prevalence of sickness absences. Such a development would be very costly for the government.

Instead, we propose the creation of directed intergovernmental grants from the state to the counties allowing MDs to spend more time with sickness absence patients. This allows the health care system to incorporate the cost of sickness absences into their decision making.

The Indirect Cost of Road Accidents

Essay 4 investigates how socioeconomic status is associated with road accidents. This study also calculates the societal costs of road accidents in terms of lost production.

Injuries due to road accidents represent a large welfare loss to society. Damage costs include a variety of expenses related to medical treatment, material and immaterial damage, law enforcement, productivity loss, and loss of time. The high costs associated with road accidents have led to heavy regulations intended to create safer transport systems and reduce the number of accidents. To that end, governments invest in road infrastructure, law enforcement and traffic education. Understanding the cost of road accidents is critical for assessing the optimal use of these instruments.

This paper uses detailed Swedish longitudinal administrative data to estimate the effect non-fatal road traffic accidents have on lost production. The costs are estimated as the present value of future earnings using an incidence approach. This means that all costs associated with the road accident are assigned to the accident year and future costs are discounted.

The first finding of this study is that the risk of being involved in a road accident is far from random. Individuals with a history of unemployment, sickness absence, social insurance uptake, and lower incomes are more likely to be involved in these accidents. These results are in accordance with previous research suggesting that involvement in road traffic injuries is associated with lower socioeconomic status (see e.g. Graham et al., 2005; Laflamme & Engström, 2002; Braver, 2003; Hasselberg et al., 2005)

Furthermore, while the utilization of sickness insurance (SI) dramatically increases at the time of the traffic accident, the effect lingers on for three years. The impact on disability insurance (DI) take-up is also large: five years after the accident the probability of receiving DI benefits has increased by around 75 percent. In economic terms, the cost of road accidents, measured as lost production, is estimated to be approximately 900 million SEK (€100 million), or 142,000 SEK per accident. For the public insurance system, the cost of road accidents is estimated to be around 410 million SEK (€45 million).

The estimated cost in this study is significantly lower than the value used by Swedish Transport Administration (STA).[3] I see two main reasons for this. (1) I am not able to account for all production losses. Only sickness absences of more than 14 days are identified. The majority of absence spells are shorter than that, and if they were included in the cost calculations, estimated costs would

[3] The Swedish Transport Administration is responsible for road maintenance and road construction and for the execution of cost-effective road construction projects. Within this framework, prospective safety improvements are given explicit monetary values. These values are then considered, together with other cost and benefits (e.g. value of travelling time and noise), in the investment appraisal.

increase. Furthermore, the cost of permanent disability among children and adolescents is not included in the estimates. (2) Previous studies have assumed that road traffic accidents are random in the population. This study shows that this is not the case. I show that if the negative selection into road traffic accidents is not taken into account, the indirect cost will be overestimated with around 19 percent.

References

Adams, P., Hurd, M. D., McFadden, D., Merrill, A. & Ribeiro, T. (2003). 'Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status'. *Journal of Econometrics*. 112(1), 3 - 56.
Antonovsky, A. (1967). 'Social class, life expectancy and overall mortality', *The Milbank Memorial Fund Quarterly*, 45(2), 31 - 73.

Braver, E. (2003). 'Race, hispanic origin, and socioeconomic status in relation to motor vehicle occupant death rates and risk factors among adults', *Accident Analysis & Prevention*, 35(3), 295 – 309.

Case, A. & Deaton, A. (2005a). 'Health and wealth among the poor: India and South Africa compared'. *American Economic Review*, 95(2), 229 - 233.

Case, A. & Deaton, A. S. (2005b). Broken down by work and sex: How our health declines. In David Wise (ed.), *Analyses in the Economics of Aging* (pp. 185 - 212). Chicago: University of Chicago Press.

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

Frijters, P., Haisken-DeNew, J. P. & Shields, M. A. (2005). 'The causal effect of income on health: Evidence from German reunification', *Journal of Health Economics*, 24(5), 997 - 1017.

García-Gómez, P., van Kippersluis H., O'Donnell, O. & van Doorslaer E. (2013). 'Longterm and spillover effects of health shocks on employment and income'. *Journal of Human Resources*, 48(4), 873 - 909.

Graham, D., Glaister, S., & Anderson, R. (2005). 'The effects of area deprivation on the incidence of child and adult pedestrian casualties in England'. *Accident Analysis & Prevention*, 37(1), 125-135.

Hasselberg, M., Vaez, M., & Laflamme, L. (2005). 'Socioeconomic aspects of the circumstances and consequences of car crashes among young adults'. *Social science & medicine*, 60(2), 287-295.

Henrekson, M. & Persson, M. (2004). 'The effects on sick leave of changes in the sickness insurance system', *Journal of Labor Economics*, 22(1), 87 - 114.

Hesselius, P. & Persson, M. (2007).*Incentive and spill-over effects of supplementary sickness compensation* (No. 2007: 16). Working Paper, IFAU-Institute for Labour Market Policy Evaluation.

Hussey, S., Hoddinott, P., Wilson, P., Dowell, J., & Barbour, R. (2004). 'Sickness certification system in the United Kingdom: Qualitative study of views of general practitioners in Scotland'. *BMJ: British Medical Journal*, 328(7431), 88.

Johansson, P. & Palme, M. (1996). 'Do economic incentives affect work absence? Empirical evidence using Swedish micro data'. *Journal of Public Economics*, 59(2), 195 - 218.

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

Johansson, P. & Palme, M. (2005). 'Moral hazard and sickness insurance', *Journal of Public Economics*, 89(9-10), 1879 - 1890.

Laflamme, L., & Engström, K. (2002). 'Socioeconomic differences in Swedish children and adolescents injured in road traffic incidents: cross sectional study'. *Bmj*, 324(7334), 396-397.

Löfgren, A., Hagberg, J., Arrelöv, B., Ponzer, S., & Alexanderson, K. (2007). 'Frequency and nature of problems associated with sickness certification tasks: A cross-sectional questionnaire study of 5455 physicians'. *Scandinavian Journal of Primary Health Care*, 25(3), 178-185.

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

Pettersson-Lidbom, P. & Thoursie, P. S. (2013). 'Temporary disability insurance and labor supply: Evidence from a natural experiment'. *Scandinavian Journal of Economics*, 115(2), 485 - 507.

Smith, J. P. (1999). 'Healthy bodies and thick wallets: The dual relation between health and economic status'. *Journal of Economic Perspectives*, 13(2), 145 - 166.

Söderberg, E., & Alexanderson, K. (2005). Sickness certificates as a basis for decisions regarding entitlement to sickness insurance benefits. *Scandinavian Journal of Public Health*, 33(4), 314-320.

Wahlström, R., & Alexanderson, K. (2004), Physicians' sick-listing practices. In K. Alexanderson & A. Norlund (eds.), *Sickness absence - causes, consequences, and physicians' certification practice: A systematic literature review by the Swedish Council on Technology Assessment in Health Care. Scand J Public Health*, 32(suppl 63), 222-255.

van Kippersluis, H., Van Ourti, T., O'Donnell, O. & van Doorslaer, E. (2009). 'Health and income across the life cycle and generations in Europe'. *Journal of Health Economics*, 28(4), 818 - 830.

Ziebarth, N. R. (2013). 'Long-term absenteeism and moral hazard—Evidence from a natural experiment'. *Labour Economics*, 24, 277-292.

Ziebarth, N. R. & Karlsson, M. (2013). 'The effects of expanding the generosity of the statutory sickness insurance system', *Journal of Applied Econometrics*.

Essay 1

Heterogeneity in the Impact of Health Shocks on Labor Outcomes: Evidence from Swedish Workers

1 Introduction

The socioeconomic gradient in health is one of the most widely replicated results in the social sciences. It dates back to at least the 19th century, where researchers have documented marked health differences across different groups in society, such as the royalty, the land-elite, and the working class.[1] In modern societies, the gradient is usually found to widen during working life but then narrows as people reach older ages.[2] As shown in Figure 1, this pattern also holds in countries with universal health insurance coverage and high quality medical care, such as Sweden.[3] The figure shows the fraction in the upper and bottom income quartile at different ages stating that they are in bad health.

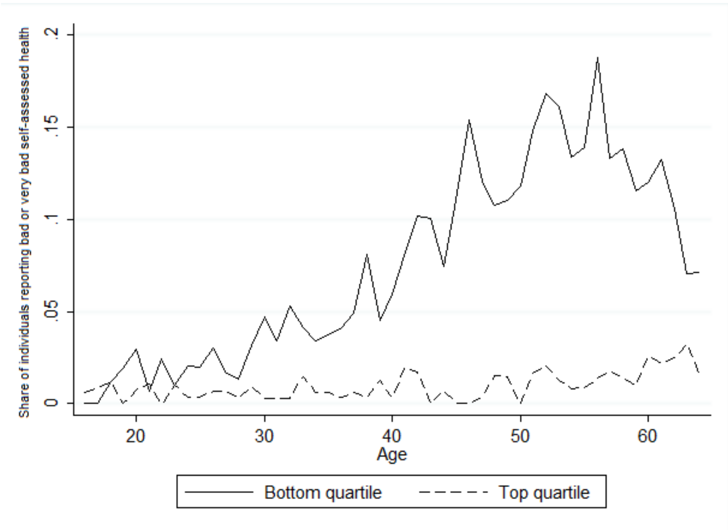


Figure 1: Share of individuals reporting bad or very bad self-assessed health by age and labor income quartile. Based on Swedish survey data from 1975 - 2007.

While there is general agreement about the existence of a socioeconomic gradient in health, there is less agreement about its underlying causes. In the epidemiological literature, it has traditionally been assumed that socioeconomic status affects health. Economists have instead explored the hypothesis that health outcomes influence socioeconomic status and surveys by Smith (2009) and Case & Deaton (2005*a*) argue that a larger part of the association between health and socioeconomic status at middle and older ages likely reflects an im-

[1]See Antonovsky (1967) for a review of the early literature on the socioeconomic gradient in health.

[2]See e.g. van Kippersluis et al. (2009), Case & Deaton (2005*b*).

[3]The survey data comes from the Swedish database ULF (Survey of living standards) It is conducted on a yearly basis and covers a random sample of about 3000 individuals.

pact of health on socioeconomic status.[4]

The recent economics literature has for the most part implicitly assumed that the impact of a health event on labor market outcomes is similar across different socioeconomic groups.[5] This is a rather restrictive assumption for a number of reasons. First, a substantial literature has shown socioeconomic heterogeneity in the recovery and survival from medical conditions, such as cancers and heart diseases (e.g. Schrijvers & Mackenbach, 1994). Results have also shown that high educated individuals are better at adhering to medical treatments, such as AIDS and diabetes treatments (Goldman & Smith, 2002). Second, a recent literature has shown differences in access to medical technologies and treatments by education (Rosvall et al., 2008). Third, people of different socioeconomic status may face different incentives to return to the labor market after facing a health shock. Fourth, the extent to which job tasks require good physical health likely varies according to socioeconomic status. Finally, it is likely that individuals of high socioeconomic status could easier change occupation or in other ways adjust their work conditions in response to a health shock. The protective role of education for health has also been emphasized by many economists, who have argued that more educated people are better at understanding and using health information and navigating the health care system (Deaton & Paxson, 2001).

If there exists substantial heterogeneity in the effect of health shocks, this could be an important part of the explanation to how the socioeconomic gradient in health arises.[6] In this paper, we advance the recent literature by focusing on heterogeneity by education in the relative effect of health shocks on labor outcomes. In addition, we will allow the effects to vary by age. This is suggested by the *cumulative advantage hypothesis*, where certain mediators of the relationship between socioeconomic status and health (e.g. smoking or social capital) accumulates over the life cycle.[7] Older individuals from lower socioeconomic groups may therefore be especially sensitive to health shocks, which partly explains why the socioeconomic gradient in health increases in the middle-ages.

[4] The conclusion of little or no effect of income on health outcomes is supported by e.g. Adams et al. (2003), Frijters et al. (2005). Other studies (such as Sullivan & von Wachter, 2009) find more substantial effects of income on health.

[5] Two exceptions are Smith (1999) and García-Gómez et al. (2013), who both consider heterogeneity in the effects of health shocks by income. Smith (1999) finds that households whose pre-illness household income is above median income face similar medical expenses but larger wealth losses than below median income households. García-Gómez et al. (2013) split their sample by income quartiles and find that low income households suffer worse labour market consequence following a health shock. To the best of our knowledge, no studies have considered how the effects vary by level of education.

[6] Note also that if the *impact* of a given health shock is stronger for people of low socioeconomic status, they would face a double penalty as they already face an increased risk of experiencing negative health shocks.

[7] See e.g. Lynch, (2003) and Willson et al., (2007).

For the purpose of our study, we use longitudinal, register-based, data on education, earnings, hospitalizations, and mortality for the entire population of Swedish workers. To estimate heterogeneity in the effect of health shocks, we formulate a Difference-in-Differences design, where we directly compare low and high-educated individuals who *both* experience the same type of health shocks.[8] We then compare the change in labor outcomes over time across these matched pairs of low and high-educated workers. This design has several advantages. First, it does not require defining control groups of individuals who do not suffer from health shocks, as we directly compare outcomes only across *treated* individuals with different education but where the distribution of health shocks is the same. Without the requirement to define control groups, an important source of unobserved heterogeneity can be discarded.

Second, if individuals anticipate their health shocks, this will not affect our estimates as long as anticipation effects are similar across low and high-educated workers facing the same shocks. This is also what our results suggest. Since our data allow us to distinguish between acute and planned hospitalizations, we are also able to focus on the impact of health shocks that are largely unexpected from the individual's point of view.

Finally, we believe that a direct comparison of low and high-educated with similar health shocks makes the analysis very transparent. In a working paper version of this paper (Lundborg et al., 2011) we have also performed analysis using more traditional control groups. Specifically, we used individuals who did not suffer a health shock in a given year to estimate the overall effect of health shocks as well as separate effects for low- and high educated. This analysis resulted in similar conclusions as in this paper.

Our data also allow us to conduct a detailed investigation on possible mechanisms that may give rise to heterogeneity in the impact of health shocks. We are able to study if health shocks affect the the uptake of social benefits such as disability insurance and sickness absence or affect the risk of unemployment. In addition, we test to what extent the heterogeneity arises from differential access to health-care and treatments, differences in the severity of health shocks, differences in occupations, and differences in the incentives to return to work after a health shock across socioeconomic groups. We also investigate if health shocks affects outcomes such as fertility, spouse's earnings, and marital status differently across socio-economic groups.

Due the richness of our data, we are able to estimate the relative importance of various types of health shocks in a given population, whereas most previous studies have focused on the impact of one particular health event at a time.[9] We

[8] Other recent studies that address the endogeneity of health shocks when studying labour market outcomes include Au et al., (2005), Disney et al. (2006), García-Gómez & Nicolás (2006), García-Gómez et al. (2013), Halla & Zweimüller (2013).

[9] See e.g. Dano (2005) on the effects of accidents in Denmark.

are also able to follow individuals for up to 12 years with essentially no attrition and to consider both the short and long-term heterogenous impact of health shocks. Moreover, we do not have to rely on self-reported measures of health, where there is substantial evidence that reporting bias differ by socioeconomic status (d'Uva et al., 2008).

Our results show that there is substantial heterogeneity in the relative impact of health shocks by education. In the short-run, the effect of a health shock is much greater for low-educated individuals. The difference is most pronounced for older individuals, where the effect for individuals with low education is more than twice that for individuals with high education.

Our results also suggest some interesting time patterns. For young individuals the difference between individuals with low and high education decreases with time. At older ages, the difference instead increases with time since the health shock. This suggests that at old ages, there are very large differences in the long-run possibilities to cope with a negative health shock. These results are consistent with the idea that the socioeconomic gradient in health is partly caused by the impact of health shocks on socioeconomic status. They also offer one explanation for why the socioeconomic gradient in health widens during middle ages.

As for mechanisms, we show that low-educated individuals are much more likely to take up disability insurance, sickness insurance, and unemployment benefits following a health shock. We find no evidence, however, that these differences are explained by differences in the economic incentives to return to the workplace, in the quality of treatment or in the severity of the health shock, and differences in sector of employment only explain a smaller part of the difference.

Besides contributing to the knowledge about the causes of the socioeconomic gradient in health, we believe that improved knowledge about heterogeneity in the impact of health shocks on labor market outcomes has important policy implications. Such knowledge may point to the possibility of targeted efforts towards groups who suffer disproportionally from health shocks. Example of such policies are more intense screening for health markers among socioeconomic risk groups, regular health check-ups, and improving adhesion to treatment. Moreover, the results may provide valuable information for evaluations of the cost effectiveness of various medical interventions designed to prevent or cure disease. Lastly, our results support the idea of a protective role of education on health.

2 Data, Sampling and Descriptive Statistics

Our data is created by merging the population register LOUISE, covering basic demographic and socio-economic information on the entire Swedish population, with the Swedish National Patient Register (NPR). The latter includes information on all episodes of in-patient care in Sweden from 1987 and onwards. It includes information such as date of admission, whether the admission was acute or planned, length of stay, as well as diagnosis (through the International Classification of Diseases, ICD). In addition, we exploit data from the National Causes of Death register.

We focus on in-patient admissions during the period 1992-2000, allowing us to observe earnings both before and after the health shock. We restrict our sample to individuals who had their first *acute* admission during this time period, since we want to focus on health shocks that are unexpected from the individual's point of view.[10] For the same reason we exclude individuals that had a planned admission before their first acute admission. We also exclude individuals that experience an admission shortly before the start of our observational period (in the period 1990-1991). We further restrict our sample to individuals who are aged 30-59 when they suffer a health shock, since a high fraction of those younger than 30 have not yet entered the labor market and many of those older than 59 are about to retire from the labor market.

We classify all the admissions into 19 major types of diseases, based on ICD-codes. Out of these we choose to focus on the ten most common (in terms of incidence).[11]

Since our focus is on the effect of health shocks on labor market outcomes, we exclude individuals that are not part of the labor force. In our main specification, we therefore only include individuals who participated in the labor force five to two years prior to the health shock. We have performed robustness analyses with respect to this restriction, and our results are insensitive to other ways of restricting the sample to those participating in the labor force. We define labor force participation as having a yearly labor income larger than one Price Base Amount (between 33,000 SEK (€3,300) and 38,000 SEK (€3,800) depending on year).[12] Note that this is a rather lax condition that essentially only removes permanently non-employed individuals.

Our main outcome is annual labor income, measuring all cash compensation paid by employers. We have data for the period 1987-2004, and use income 5

[10] After suffering a first health shock, some individuals might anticipate additional shocks. This is why we only focus on the first acute admission. Also, some acute health shocks may be more or less expected, such as those resulting from heavy smoking and drinking, but their exact timing is usually unknown in advance.

[11] Using ICD-9/10 codes, these were classified as follows: infectious diseases (0010-139; A00-B99), cancer (140-239; C00-D48), mental and behavioral diseases (290-319; F00-F99), nerve system (320-359; G00-G99), respiratory diseases (460-519; J00-J99), heart (390-459; I00-I99), digestive organs (520-579; K00-K93), musculoskeletal system and connective tissues (710-739; M00-M99), genitourinary system (580-629; N00-N99) and external accidents (800-1000; S00-T98).

[12] The price base amount is a measure calculated by the government each year, based on changes in the consumer price index.

years before the shock and up to 10 years after the shock. In order to measure relative effects we use log earnings as outcome.[13] From the registers we also have information on income from various social insurance schemes, spousal income, marital status and number of children for the period 1990-2004.

Table 1 provides descriptive statistics on the fraction of the working population affected by a health shock by age and level of education in 1995. Health shocks become more common as people reach old age and there are large differences by level of education. About 3.5 percent experience a health shock in the youngest age group (30-39), whereas the corresponding figure in the oldest age group (50–59) is 5.5 percent. Individuals without a university education are much more likely to be affected by negative health shocks compared to individuals with a university education. For instance, in the youngest age group, the share with a health shock is about 40 percent larger among the low educated group compared to the group with high education. A notable exception is cancer, for which the incidence is the same regardless of educational background.

[13] Individuals with zero yearly labor income are assigned 6000 SEK (€600). This basically corresponds to the lowest yearly income observed in our data.

Table 1: Sample statistics for health shocks

	Age 30-39		Age 40-49		Age 50-59	
	Low edu.	High edu.	Low edu.	High edu.	Low edu.	High edu.
Any Shock (%)	3.90	2.46	4.50	2.94	5.69	4.12
Infectious diseases (%)	0.19	0.17	0.16	0.14	0.19	0.16
Cancer (%)	0.08	0.08	0.18	0.18	0.30	0.29
Mental & behavioral (%)	0.79	0.27	0.91	0.36	0.73	0.39
Nerve system (%)	0.12	0.06	0.13	0.08	0.15	0.10
Heart diseases (%)	0.17	0.12	0.48	0.30	1.18	0.77
Respiratory diseases (%)	0.25	0.16	0.24	0.17	0.37	0.23
Digestive organs (%)	0.46	0.34	0.57	0.38	0.73	0.53
Musculoskeletal (%)	0.22	0.13	0.27	0.18	0.33	0.22
Genitourinary (%)	0.29	0.19	0.30	0.23	0.30	0.25
External accidents (%)	0.74	0.39	0.75	0.47	0.79	0.60

Notes: The table reports the fraction affected by any health shocks and the ten most common types of health shocks in 1995. High education is defined as having some kind of university education and low education less than university education.

3 Empirical Strategy

In order to estimate heterogeneity in the impact of acute health shocks by level of education we directly compare the responses of individuals with a university education (high education) to those of individuals without a university education (low education). We have also performed analyses using finer educational groups, resulting in similar conclusions. This approach allows us to estimate heterogeneity in the effects without having to define control groups of individuals not experiencing any health shocks.

For each low-educated person, we use nearest neighbor propensity score matching to match a high-educated person that experience the same type of health shock (10 categories) in the same calender year.[14] We further match on gender, age in months at the time of the health shock and the quarter when the shock occurred. This controls for important differences between high and low educated while simultaneously avoiding matching on covariates, such as sector of employment and marital status, that are partly determined by level of edu-

[14]We have also started from the high-educated and searched for matches among the low-

cation. In our robustness analysis, we will check to what extent the results are affected by matching on some additional covariates.

Besides these observed differences low- and high-educated with the same type of health shock might differ in other ways, such as unobserved abilities and underlying health. For that reason the matched sample is used in a difference-in-differences analysis using outcome data both before and after the health shock. In this analysis the changes in income between before and after shock for low-educated are contrasted with the same change for high-educated. Thus, if the outcomes change more between before and after the health shock for low-educated compared to high-educated we interpret this as evidence of heterogenous effects. This controls for all fixed differences between low- and high-educated that are unrelated to the differential impact of health shocks. Specifically, our baseline heterogenous effects model, for log labor earnings for individual i at t years after the health shock is

$$\log(y_{it}) = \lambda_i + \lambda_t + \sum_{\tau=0}^{T} \delta_\tau^{LE} I(t=\tau) I^{LE} + \gamma_1 I^{LE} t + \gamma_2 (1 - I^{LE}) t + \varepsilon_{it}. \quad (1)$$

In the shock year $t = 0$ and $t > 0$ ($t < 0$) for the years after (before) the shock. I^{LE} is an indicator variable taking the value one if the individual has low education. The coefficients $\delta_0^{LE}, ..., \delta_T^{LE}$ capture the immediate effects of the health shock, the effect on year after the shock, and so on. The model includes individual fixed effects, λ_i and time fixed effects, λ_t. The former control for all time-invariant factors at the individual level, such as labor preferences, early life environment and/or underlying ability. The latter capture general changes over time, and for the years after the health shock, also the "effects" for the high-educated. We also control for underlying pre-shock trends in labor earnings by including separate linear trends for low- and high-educated.[15] In the analyses we also run separate regressions for three different age groups.[16]

After controlling for fixed effects and general pre-shock trends, there may remain some pre-treatment effects since health shocks could still be more or less anticipated and/or show an effect already before the actual admission. Here, focusing on heterogenous effects helps. If the anticipation effects are similar for individuals with high and low education, our heterogenous effects estimate could still be given a causal interpretation. In Section 4.4 we shed more light

educated. This analysis resulted in very similar results.

[15] We have also run models using quadratic trends but the results were insensitive to including more flexible controls for trends.

[16] Note that this two-step procedure ignores uncertainty in the matching step. For that reason we have validated our results by performing genuine conditional Difference-in-Differences, using nearest neighbor mahalanobis-metric matching and the Abadie & Imbens (2006) estimator of the standard errors. From this exercise we concluded that taking uncertainty in the matching step into account has little impact on our inference.

on such anticipation effects and perform a variety of other robustness checks on our data. This includes running placebo estimates to test for any significant differences in pre-shock responses. We also use detailed data on the number of medical procedures and number of diagnoses in order to investigate whether our results are driven by differences in the severity of the health shocks. In addition, we examine if differential death rates by level of education affect our estimates.

4 Main Results

4.1 Graphical Evidence

We start by illustrating some of the most important patterns in our data graphically. Figure 2 shows the average log labor earnings for the matched sample of low- and high-educated, and Figure 3 displays the difference between these two averages, adjusting for the difference five years before the shock. Pre-shock earnings are higher among the high educated and also increase faster among the high educated. There is also a tendency for earnings to flatten out one year before the health shock but this pattern is equally strong for low- and high-educated. These patterns illustrate the importance of taking fixed unobserved heterogeneity and pre-shock trends into account and also show that focusing on heterogeneous responses largely mitigates problems with anticipation effects.

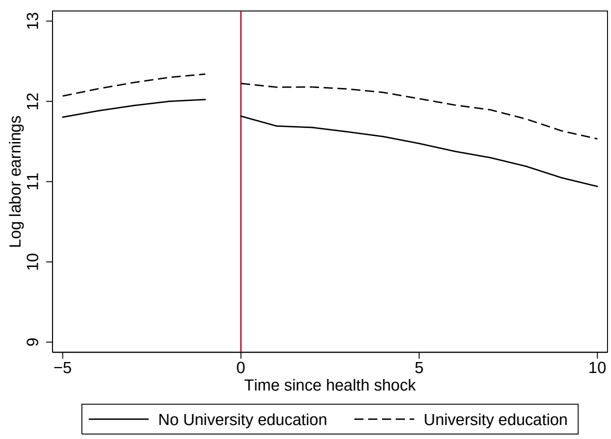


Figure 2: Log labor earnings for individuals with and without university education

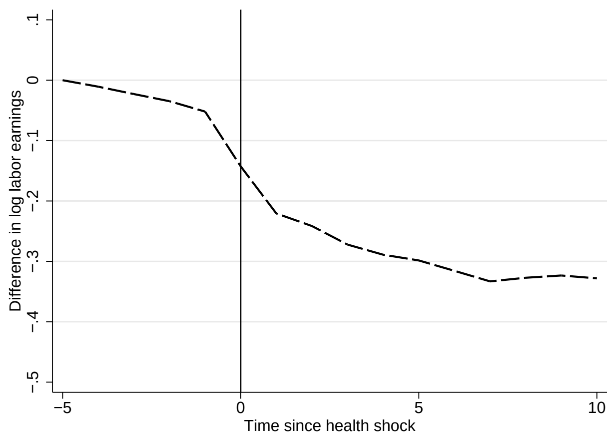


Figure 3: Difference in average log labor earnings between individuals with and without university education

Figures 2 and 3 also reveal heterogeneity in the effects. In the year of the health shock, income decreases relatively more among low-educated individuals. Besides this immediate difference, the figure also provide a first indication of substantial heterogeneity in the long-run effects. Several years after the health shock, income has still decreased much more among the low-educated.

4.2 Heterogeneous Effects by Level of Education

Turning to our main regression results, shown in column 1 of Table 2, labor earnings decline 8 percent more for low-educated compared to high-educated in the year of the health shock. Note that we do not correct for the fact that on average, the health shock occurs in the middle of the year, so that the income loss per month is substantially larger than 1/12 of the estimates in the table. One year later, the effect increases to 14.5 percent. This difference also persists in the long-run. For instance, eight years after the health shock the effect is about 16.2 percent larger for low-educated individuals. These results suggest that individuals from lower socioeconomic groups not only suffer from more frequent health shocks but also suffer disproportionally from a given health shock.

4.3 Education, Age, Gender and Type of Health Shock

We next run separate regressions for the three age groups; 30–39, 40–49 and 50–59. The estimates presented in columns 2-4 of Table 2 show important and significant short-term differences in the effects between low-educated and high-educated workers in all three age groups. One year after the shock, the largest heterogeneity is found in the 50–59 age group (16.6 percent) and the

Table 2: Estimates of heterogeneous effects of health shocks by level of education and age

	(1) All	(2) Age 30-39	(3) Age 40-49	(4) Age 50-59
Shock year	-0.0802**	-0.0879**	-0.0722**	-0.0823**
	(0.00414)	(0.00852)	(0.00634)	(0.00689)
Shock year+1	-0.145**	-0.134**	-0.129**	-0.166**
	(0.00622)	(0.0122)	(0.00941)	(0.0106)
Shock year+2	-0.153**	-0.135**	-0.128**	-0.185**
	(0.00735)	(0.0138)	(0.0110)	(0.0129)
Shock year+3	-0.171**	-0.137**	-0.144**	-0.215**
	(0.00843)	(0.0153)	(0.0122)	(0.0150)
Shock year+4	-0.175**	-0.127**	-0.158**	-0.219**
	(0.00956)	(0.0168)	(0.0132)	(0.0175)
Shock year+5	-0.172**	-0.123**	-0.162**	-0.212**
	(0.0111)	(0.0188)	(0.0149)	(0.0206)
Shock year+6	-0.176**	-0.124**	-0.184**	-0.205**
	(0.0125)	(0.0204)	(0.0162)	(0.0232)
Shock year+7	-0.181**	-0.110**	-0.191**	-0.224**
	(0.0139)	(0.0222)	(0.0178)	(0.0258)
Shock year+8	-0.162**	-0.0846**	-0.188**	-0.190**
	(0.0158)	(0.0255)	(0.0198)	(0.0287)
Shock year+9	-0.146**	-0.0698*	-0.186**	-0.161**
	(0.0183)	(0.0290)	(0.0227)	(0.0318)
Shock year+10	-0.138**	-0.0609	-0.193**	-0.138**
	(0.0208)	(0.0322)	(0.0256)	(0.0353)
# observations	5,544,058	1,319,757	1,947,779	2,276,522

Note: FE estimates of heterogenous effects comparing low- and high-educated using the pre-matched sample described in the data section. Outcome is log yearly labor earnings. The models also include time fixed effects and separate linear trends for individuals with and without university education. Robust standard errors in parentheses. * and ** indicate significance at 5 and 1 percent levels.

smallest in the 40–49 age group (12.9 percent). Beside these short-term differences, there are also some interesting time patterns. For the youngest group, the heterogeneity decreases with time after the shock. One year after the shock, the effect is 13.5 percent larger among low-educated, whereas eight years after the shock this effect has decreased to 8.9 percent. For individuals at middle and old ages, the additional effect for the low-educated instead gets larger for some time after the health shock. Seven year after the health shock, for instance, the effect has increased to 22.4 percent.

In Tables A:1 and A:2 in the appendix we examine to what extent the observed heterogeneity exists across different types of major health shocks. This is the case for most types of health shocks and the magnitude of the effects is rather similar across types. We have also examined if the heterogeneity exists both among males and females. The results presented in Table 3 show this to be the case. In the shock year the effect for females (males) with low education is 6.8 (8.8) percent larger compared to other females (males). Eight years later these heterogenous effects are 22.8 percent for females and 16.3 for males, so that in the long run the heterogeneity is somewhat stronger among females.

Table 3: Estimates of heterogeneous effects by level of education and gender

	(1) Men	(2) Women
Shock year	-0.0981**	-0.0677**
	(0.00644)	(0.00619)
Shock year+1	-0.171**	-0.134**
	(0.00989)	(0.00946)
Shock year+2	-0.171**	-0.165**
	(0.0123)	(0.0114)
Shock year+3	-0.188**	-0.197**
	(0.0145)	(0.0135)
Shock year+4	-0.187**	-0.220**
	(0.0168)	(0.0157)
Shock year+5	-0.187**	-0.222**
	(0.0199)	(0.0186)
Shock year+6	-0.195**	-0.231**
	(0.0227)	(0.0213)
Shock year+7	-0.208**	-0.235**
	(0.0254)	(0.0240)
Shock year+8	-0.188**	-0.226**
	(0.0288)	(0.0272)
Shock year+9	-0.173**	-0.218**
	(0.0334)	(0.0319)
Shock year+10	-0.170**	-0.212**
	(0.0384)	(0.0369)
# observations	3,125,419	2,075,807

Note: Note: FE estimates using the pre-matched sample described in the data section. Outcome is log yearly labor earnings. The models also include time fixed effects and separate linear trends for individuals with and without university education. Robust standard errors in parentheses. * and ** indicate significance at 5 and 1 percent levels.

4.4 Robustness Analysis

We next provide placebo estimates, in which we move the time of the shock one and two years back in time, that is before the actual acute admission took place. If the placebo effects differ by education, this might reflect that low-educated individuals anticipate or react to their future health shocks to a larger extent.

Column 1 of Table 4 displays the placebo estimates for the full sample and columns 2-4 estimates for the three age groups. Note that the placebo estimates should be interpreted as the *difference* in placebo effect between low and high-educated. For the all age groups combined, we find insignificant and small placebo estimates both one and two years before the shock. For the age group 40-49, however, we find a small but significant difference between low and high-educated one and two years before the shock. Note that these differences in placebo estimates are very small and the statistical significance is mainly due to the large sample size. Also, if anticipation effects are somewhat larger for low-educated individuals, it serves to attenuate the true difference in the post shock effect between low and high-educated. A small difference in placebo effect is also found for the 30-39 age group two years before the health shock. In column 4, as a comparison, we show the corresponding estimates for *planned* admissions, using the same sampling and estimation scheme as for the acute admissions. As expected, we find significant placebo estimates one year before the planned admission, and this lends support for using only unexpected acute admissions.

Next, we test whether using an extended set of covariates in the matching step affects the results. In column 1 of Table 7, results are shown when also matching on the number of children, marital status, and municipality of residence. Doing so only slightly reduces the magnitude of some of the estimates.

An alternative explanation of the larger effects for individuals with low education is that they experience more *intense* shocks, even though the diagnosis is the same. We test for this by using information on the number of diagnoses and number of medical procedures as proxies for the severity of the shock. Needless to say, this is not a perfect measure, but in Table 5 we show that the number of diagnoses is highly correlated with the fraction that dies within five years both for low- and high-educated persons. For both low- and high-educated the death rate is almost twice as high for individuals with at least two secondary diagnoses compared to individuals without a secondary diagnosis.[17] Specifically,

[17] The relation between the death rate and number of medical procedures is less clear, but the highest death rate is found for individuals with two or more procedures. For medical procedures, the interpretation is somewhat more difficult, since a greater number of medical procedures could be a sign of worse health but it could also indicate better treatment, which potentially could improve the long term outcomes.

Table 4: Placebo heterogenous effects regressions

	(1) Age 30 - 39	(2) Age 40 - 49	(3) Age 50 - 59	(4) Planned admissions
Shock year-2	0.0130*	-0.0130**	0.00183	0.000228
	(0.00581)	(0.00379)	(0.00344)	(0.00255)
Shock year-1	0.00903	-0.0210**	-0.00300	-0.00802*
	(0.00889)	(0.00595)	(0.00522)	(0.00398)
Shock year	-0.0741**	-0.0956**	-0.0838**	-0.101**
	(0.0138)	(0.00953)	(0.00962)	(0.00676)
Shock year+1	-0.117**	-0.157**	-0.167**	-0.140**
	(0.0175)	(0.0127)	(0.0132)	(0.00863)
Shock year+2	-0.115**	-0.162**	-0.187**	-0.144**
	(0.0202)	(0.0148)	(0.0160)	(0.0103)
Shock year+3	-0.114**	-0.184**	-0.217**	-0.157**
	(0.0229)	(0.0165)	(0.0183)	(0.0118)
Shock year+4	-0.101**	-0.203**	-0.222**	-0.176**
	(0.0253)	(0.0182)	(0.0210)	(0.0135)
Shock year+5	-0.0941**	-0.213**	-0.215**	-0.184**
	(0.0283)	(0.0204)	(0.0243)	(0.0153)
Shock year+6	-0.0918**	-0.241**	-0.209**	-0.185**
	(0.0311)	(0.0223)	(0.0271)	(0.0173)
Shock year+7	-0.0746*	-0.253**	-0.228**	-0.191**
	(0.0343)	(0.0245)	(0.0301)	(0.0191)
Shock year+8	-0.0460	-0.255**	-0.195**	-0.194**
	(0.0381)	(0.0270)	(0.0334)	(0.0213)
Shock year+9	-0.0281	-0.259**	-0.166**	-0.184**
	(0.0421)	(0.0302)	(0.0366)	(0.0239)
Shock year+10	-0.0161	-0.271**	-0.144**	-0.177**
	(0.0458)	(0.0332)	(0.0402)	(0.0268)
# observations	1,319,757	1,947,779	2,276,522	4,185,617

Note: Outcome is log yearly labor earnings. Columns 1-3 report FE placebo estimates which are created by artificially moving back the treatment two years, and using the pre-matched sample described in the data section. Column 4 reports similar estimates for planned admissions. Robust standard errors in parentheses. * and ** indicate significance at 5 and 1 percent levels.

we include number of diagnoses (1,2, 3, or ≥ 4) and number of procedures (0, 1, 2, or ≥ 3) in our set of covariates in the matching step and estimate model (1) using this new set of matched low- and high-educated. We obtain results very similar to our main results (column 2 of Table 7). From this we conclude

Table 5: Death rates by number of diagnoses and medical procedures

	Age 30 - 39		Age 40 - 49		Age 50 - 59	
	Low edu.	High edu.	Low edu.	High edu.	Low edu.	High edu.
# of diagnoses						
1	0.95	0.58	2.25	1.75	4.77	3.28
2	1.52	0.69	3.24	2.80	6.07	5.17
3	1.82	1.27	4.13	3.03	8.69	6.35
≥ 4	3.22	4.90	4.65	4.58	11.17	8.92
# of medical procedures						
0	1.32	0.71	2.84	2.29	5.55	4.06
1	0.66	0.44	1.74	1.58	4.29	3.46
2	0.77	1.11	2.35	1.36	5.45	3.30
≥ 2	1.84	1.15	4.28	3.82	9.36	6.21
# observations	47,648	47,648	69,848	69,848	84,464	84,464

Note: The table reports the fraction in the analysis sample who dies within five year from the shock year. Number of diagnoses are number of secondary diagnoses including the main diagnosis. Number of medical procedures counts registered medical procedures for the current hospitalization. High education is defined as having some kind of university education and low education less than university education.

that even if these measures are imperfect proxies for the severity of the health shock it is unlikely that the observed heterogeneous effects purely are an effect of differences in the severity of the health shocks.

If the fraction that dies during our observation window differ systematically by level of education our estimates may be biased. In Table 6 we show that low educated are indeed more likely to die. This may lead us to underestimate the true difference in effects across low and high-educated groups, since the fraction of "frail" individuals in the low-educated group will decrease faster over time. However, re-estimating our main model, but now only including matched pairs that survive the *entire* observation period, produces results very similar to our main results (column 3 in Table 7).

Finally, our estimates are based on an unbalanced panel, because we are not able to follow individuals who suffered their health shock quite recently for a full ten years after the health shock. As shown in column 4 in Table 7 , including only individuals that are observed all ten years after the health shock do not change the results to any important extent.[18]

[18]Another issue concerns differences in health-seeking behaviour between low and high-educated individuals. If high-educated individuals are better at seeking inpatient care when needed, the composition of the groups is affected. The low-educated individuals in our sample would then to a greater extent be positively selected, which might serve to downward bias the difference in effects between low and high-educated persons. We do not believe this issue to be of any great importance, since the health shocks we consider are of acute character and require hospital care in most cases. Moreover, the patient fee for obtaining treatment is very low as hospitalizations are covered by the public health insurance.

Table 6: Death rates by level of education, age and type of health shocks

	Age 30 - 39		Age 40 - 49		Age 50 - 59	
	Low edu.	High edu.	Low edu.	High edu.	Low edu.	High edu.
Total	1.11	0.67	2.56	2.04	5.40	3.92
Infectious	0.97	1.01	1.81	1.01	2.81	1.01
Cancer	12.51	21.81	16.35	21.81	32.56	21.81
Mental	2.65	3.33	5.06	3.33	8.25	3.33
Nerve	1.32	2.13	2.38	2.13	5.33	2.13
Heart	1.23	2.66	2.26	2.66	4.55	2.66
Respiratory	0.68	1.92	2.28	1.92	5.74	1.92
Digestive	0.67	1.47	1.47	1.47	3.51	1.47
Musculoskeletal	0.28	1.13	0.98	1.13	1.99	1.13
Genitourinary	0.43	0.90	0.84	0.90	2.39	0.90
External	0.62	1.26	1.29	1.26	2.63	1.26
# observations	47,648	47,648	69,848	69,848	84,464	84,464

Note: The table reports the fraction in the analysis sample who die within 5 years of the shock year. High education is defined as having some kind of university education and low education less than university education.

Table 7: Robustness analysis

	(1) Extended	(2) Severity	(3) Survivors	(4) Balanced panel
Shock year	-0.0704**	-0.0755**	-0.0738**	-0.0857**
	(0.00384)	(0.00349)	(0.00421)	(0.00542)
Shock year+1	-0.131**	-0.136**	-0.134**	-0.141**
	(0.00613)	(0.00535)	(0.00624)	(0.00813)
Shock year+2	-0.140**	-0.147**	-0.138**	-0.148**
	(0.00691)	(0.00633)	(0.00741)	(0.00954)
Shock year+3	-0.152**	-0.155**	-0.153**	-0.165**
	(0.00787)	(0.00730)	(0.00848)	(0.0108)
Shock year+4	-0.153**	-0.159**	-0.157**	-0.170**
	(0.00890)	(0.00830)	(0.00960)	(0.0122)
Shock year+5	-0.152**	-0.159**	-0.153**	-0.166**
	(0.0102)	(0.00954)	(0.0112)	(0.0140)
Shock year+6	-0.150**	-0.156**	-0.151**	-0.163**
	(0.0115)	(0.0108)	(0.0126)	(0.0152)
Shock year+7	-0.148**	-0.152**	-0.157**	-0.170**
	(0.0128)	(0.0121)	(0.0140)	(0.0162)
Shock year+8	-0.130**	-0.143**	-0.134**	-0.151**
	(0.0147)	(0.0138)	(0.0159)	(0.0173)
Shock year+9	-0.112**	-0.127**	-0.111**	
	(0.0170)	(0.0158)	(0.0185)	
Shock year+10	-0.107**	-0.122**	-0.100**	
	(0.0192)	(0.0180)	(0.0210)	
# observations	5,386,780	5,542,053	5,233,694	3,481,846

Note: Outcome is log yearly labor earnings. Column 1 includes municipality dummies, number of children and marital status as additional covariates in the matching step. Column 2 includes measures of severity of the shock. Column 3 restricts the sample to matched pairs where both survives the entire observation period and Column 4 is based on a balanced panel. All models also include time fixed effects and separate linear trends for individuals with and without university education. Robust standard errors in parentheses. * and ** indicate significance at 5 and 1 percent levels.

5 Mechanisms

5.1 Labor Supply and Uptake of Social Benefits

In order to better understand the mechanisms through which the heterogeneity arises, we next conduct a detailed investigation of what *types* of labor market outcomes individuals experience after a health shock. Individuals may, for instance, retire entirely from the labor market, decrease their working time, stay on sickness absence or disability insurance benefits, or become unemployed. Such information could, for instance, be used to design policies with the aim of mitigating the effects of health shocks.

The first column of Table 8, shows that low-educated are more likely to leave the labor market entirely after experiencing a health shock, both in the short and in the long run. Six years after the health shock, the fraction of un-employed is almost 7 percentage points higher among the low-educated. In the short run this is mainly explained by increased sickness absence rates. Long run differences between low and high educated are mainly explained by higher disability rates, but also to some extent by higher unemployment. For instance, the low educated experience a 6 percentage points higher disability rate and a 1 percentage point higher unemployment rate six years after shock.[19]

5.2 Family Decisions

A health shock may also impact marital decisions, fertility decisions, as well as the spouse's productivity and labor supply. We next investigate if there is heterogeneity across these dimensions as well as between low and high educated persons. Column 5 of Table 8 shows that high-educated are *more* likely to experience a marital breakdown following a health shock. A similar pattern is found for fertility, where a health shock impacts more heavily on a high-educated persons propensity to have additional children. These results are consistent with high-educated person's having better outside options on the marriage market, where the spouse is more inclined to leave the marriage as the quality of the match goes down.

The effect on spouse's earnings is theoretically ambiguous. The spouse might compensate for the income loss in the family by increasing their own labour supply, but labour supply could also go down if the spouse need to care for the one affected by the health shock. Interestingly, we find no difference in the effect of a health shock on the spouse's earnings. Results not reported in the paper show that this holds for both male and female spouses. This suggests the that the compensating effect and caring effect are of roughly the same size for both males and females.

[19]We have also tested for wealth effects using information on yearly capital income, but found no significant effects.

Table 8: Heterogenous effects of health shocks and effects for income components

	(1) Employment	(2) Sickness absence	(3) Unemployment	(4) Disability insurance	(5) Marital status	(6) Fertility	(7) Spousal income
Shock year	-0.0207**	0.0499**	0.0154**	0.00173*	-0.00203	-0.000737	-0.0128
	(0.00151)	(0.00418)	(0.00166)	(0.000791)	(0.00158)	(0.00165)	(0.00771)
Shock year+1	-0.0428**	0.0416**	0.0194**	0.0105**	-0.000823	0.00149	-0.0178
	(0.00229)	(0.00442)	(0.00236)	(0.00137)	(0.00246)	(0.00183)	(0.0113)
Shock year+2	-0.0500**	0.0287**	0.0203**	0.0235**	0.00123	0.00378	-0.0180
	(0.00284)	(0.00473)	(0.00294)	(0.00193)	(0.00321)	(0.00222)	(0.0144)
Shock year+3	-0.0565**	0.0158**	0.0190**	0.0339**	0.00512	0.00445	-0.00285
	(0.0177)	(0.00546)	(0.00349)	(0.00233)	(0.00392)	(0.00258)	(0.000660)
Shock year+4	-0.0620**	0.00832	0.0149**	0.0412**	0.00761	0.00620*	0.000440
	(0.00384)	(0.00622)	(0.00402)	(0.00270)	(0.00465)	(0.00302)	(0.0211)
Shock year+5	-0.0631**	0.00166	0.0131**	0.0489**	0.0111*	0.00818*	0.0173
	(0.00443)	(0.00703)	(0.00456)	(0.00304)	(0.00542)	(0.00342)	(0.0243)
Shock year+6	-0.0683**	-0.00284	0.0113*	0.0504**	0.0136*	0.00796*	0162
	(0.00499)	(0.00785)	(0.00512)	(0.00338)	(0.00621)	(0.00386)	(0.0278)
Shock year+7	-0.0705**	-0.00495	0.00593	0.0555**	0.0166*	0.00891*	0.0331
	(0.00558)	(0.00868)	(0.00563)	(0.00367)	(0.00700)	(0.00428)	(0.0316)
Shock year+8	-0.0685**	-0.00689	-0.00354	0.0584**	0.0190*	0.0106*	0.0407
	(0.00626)	(0.00953)	(0.00619)	(0.00411)	(0.00782)	(0.00467)	(0.0356)
Shock year+9	-0.0721**	-0.0136	-0.0138*	0.0603**	0.0228**	0.0124*	0.0588
	(0.00699)	(0.0104)	(0.00670)	(0.00440)	(0.00865)	(0.00508)	(0.0406)
Shock year+10	-0.0752**	-0.0203	-0.0256**	0.0615**	0.0237*	0.0109	0.0923*
	(0.00785)	(0.0113)	(0.00725)	(0.00491)	(0.00956)	(0.00557)	(0.0458)
Means	0.87	0.14	0.07	0.02	0.61	0.02	188,296
# observations	5,327,578	4,734,816	5,201,229	5,201,229	5,199,151	5,327,578	3,050,059

Note: FE estimates using the pre-matched sample described in the data section. Outcomes in Columns 1-4 are indicators for employment and take-up of sickness insurance, disability insurance, unemployment insurance and age pension. Outcomes in Columns 5-6 are indicator for being married, indicator for at least on child aged 0-3 and log yearly spousal income. For sickness absence we use data for 1992 - 2004,and for all other outcomes data over 1990 - 2004. All models also include time fixed effects and separate linear trends for individuals with and without university education. Robust standard errors in parentheses. * and ** indicate significance at 5 and 1 percent levels.

5.3 Incentives, Treatment Quality, and Workplace

There are several other possible explanations for the observed heterogeneity. First, the economic incentive to return to work after a health shock is stronger among the high educated, due to their higher earnings and the fact that the maximum benefit levels in the Swedish social insurance are capped at a relatively low ceiling. The latter means that the effective replacement rate decreases with income.[20] If differences in economic incentives are an important factor, we expect individuals with high education to leave the hospital more quickly and return to work. We test this hypothesis by controlling for the length of the stay at the hospital. Specifically, we divide the number of days into 1-5 days, 6-10 days, 11-15 days, 16-20 days and $\geq$21 days, and interact dummy variables for each of these groups with time since the health shock.

Second, a high educated person might be better at handling contacts with the health care system, leading to higher quality treatment. We test for this by including an indicator for whether the treating hospital was a university hospital. In Sweden, university hospitals have the most advanced medical technology. It is therefore not farfetched to assume that university hospital status provides a quite good measure of the quality of the treatment.

Third, low educated are to a higher extent employed in blue collar occupations, for which the return to work after a decline in health is more complicated compared to white collar professions. We examine this hypothesis by including the individual's sector of employment (2 digit level) the year prior to the shock in the matching step. This assures that the fraction from each sector is the same among both the low- and high-educated.

The results from these different estimations are presented in columns 2-4 in Table 9. These new estimates are in most cases similar to the baseline estimates, both in the short- and long-run. The only exception is that the heterogeneous effects are somewhat smaller when controlling for sector of employment. Thus, we obtain some evidence that differences in sector of employment explain the observed heterogeneity, but no evidence that different moral hazard profiles and differences in treatment quality are important.

[20] The Swedish sickness insurance provides economic compensation when a worker is too sick to carry out his or her regular job. This insurance automatically covers all of the employed workers. The benefits in the Swedish sickness insurance are income related and the size of the benefits depends on the person's wage prior to the sick spell. The insurance consists of two main benefits: sickness benefit (SI) and disability benefit (DI). The SI is supposed to cover part of the income loss due to temporary illness. DI compensates individuals whose work capacity is permanently reduced. The replacement rates have changed over time, but the rates have been capped at a relatively low ceiling throughout our observed time period. (About 25 percent of the workers have income above the ceiling).

Table 9: Estimates of heterogenous effects controlling for length of stay, type of hospital and occupation

	(1) Baseline	(2) Length of stay	(3) University hospital	(4) Occupation
Shock year	-0.0802**	-0.0774**	-0.0806**	-0.0615**
	(0.00414)	(0.00412)	(0.00417)	(0.00442)
Shock year+1	-0.145**	-0.141**	-0.146**	-0.123**
	(0.00622)	(0.00617)	(0.00623)	(0.00703)
Shock year+2	-0.153**	-0.150**	-0.154**	-0.116**
	(0.00735)	(0.00733)	(0.00738)	(0.00850)
Shock year+3	-0.171**	-0.167**	-0.172**	-0.129**
	(0.00843)	(0.00840)	(0.00845)	(0.00975)
Shock year+4	-0.175**	-0.171**	-0.175**	-0.138**
	(0.00956)	(0.00953)	(0.00958)	(0.0108)
Shock year+5	-0.172**	-0.167**	-0.172**	-0.142**
	(0.0111)	(0.0111)	(0.0112)	(0.0123)
Shock year+6	-0.176**	-0.171**	-0.176**	-0.141**
	(0.0125)	(0.0125)	(0.0126)	(0.0139)
Shock year+7	-0.181**	-0.176**	-0.181**	-0.131**
	(0.0139)	(0.0139)	(0.0139)	(0.0158)
Shock year+8	-0.162**	-0.157**	-0.162**	-0.121**
	(0.0158)	(0.0157)	(0.0158)	(0.0178)
Shock year+9	-0.146**	-0.139**	-0.146**	-0.115**
	(0.0183)	(0.0182)	(0.0183)	(0.0205)
Shock year+10	-0.138**	-0.132**	-0.138**	-0.106**
	(0.0208)	(0.0208)	(0.0208)	(0.0233)
# observations	5,544,058	5,544,058	5,544,058	5,496,654

Note: FE estimates using the pre-matched sample described in the data section. Outcome is log yearly labor earnings. Column 1 reports our main estimates, and columns 2 and 4 include controls for the length of stay in the hospital (1-5 days, 6-10 days, 11-15 days, 16-20 days, and $\geq$ 21 days), and admission to university hospital, respectively. Column 3 has sector of employment (two digit level) included in the matching step. The models also include time fixed effects and separate linear trends for individuals with and without university education. Robust standard errors in parentheses. * and ** indicate significance at 5 and 1 percent levels.

5.4 Subsequent Health Shocks

Figure 4 shows that a substantial fraction of individuals with an initial shock experience additional shocks, and that this fraction is higher among individuals with low education. The long-term effects that we observe could, thus, either be due to long-lasting effects of the initial shock or due to these subsequent health shocks. Although correcting for selection into additional shocks is problematic, we can provide some suggestive evidence by simply adding controls for subsequent shocks into our main model.[21] Initially we control for having a second shock and stepwise allow the effect to vary by level of education and time since the second shock. Later we also control for a third health shock.

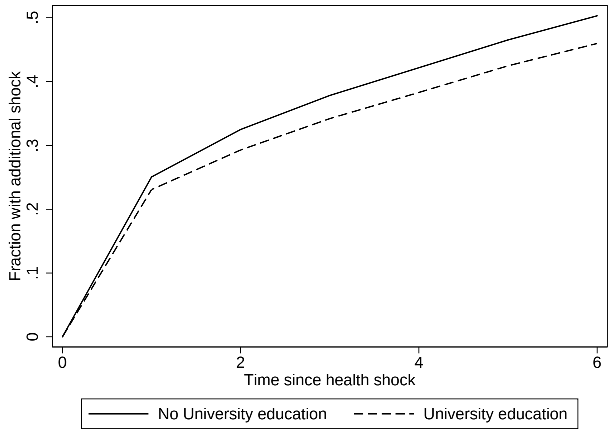


Figure 4: Fraction with a second health shock by level of education

The results from this exercise, presented in Table 10, show that both the short- and long-run heterogeneity decrease somewhat as more controls for subsequent shocks are added to the model. In our most extended model the effect in the shock year decreases from 8 percent to 7.9 percent and the effect seven years after the shock decreases from 18.1 percent to 15.3 percent. We conclude that the initial shock in itself has long-lasting impact, changing the trajectory of income for individuals with low education and that controlling for additional health shocks only has minor impact on the results.

[21]Note that we include controls that are affected by the size of the effects of the initial shock. The regression estimates, thus, only provide suggestive evidence.

Table 10: Estimates of heterogeneous effects of health shocks by level of education with controls for subsequent health shocks

	(1) Basic	(2) +2^{nd} shock	(3) +2^{nd} $*$ low edu.	(4) +time var. impact 2^{nd}	(5) +3^{rd} shock
Shock year	-0.0756**	-0.0712**	-0.0607**	-0.0622**	-0.0613**
	(0.00414)	(0.00415)	(0.00463)	(0.00431)	(0.00459)
Shock year+1	-0.143**	-0.133**	-0.116**	-0.119**	-0.118**
	(0.00623)	(0.00620)	(0.00677)	(0.00658)	(0.00675)
Shock year+2	-0.149**	-0.136**	-0.116**	-0.118**	-0.116**
	(0.00752)	(0.00751)	(0.00802)	(0.00807)	(0.00802)
Shock year+3	-0.165**	-0.151**	-0.128**	-0.129**	-0.126**
	(0.00858)	(0.00855)	(0.00912)	(0.00931)	(0.00913)
Shock year+4	-0.172**	-0.158**	-0.132**	-0.134**	-0.128**
	(0.00968)	(0.00964)	(0.0102)	(0.0105)	(0.0102)
Shock year+5	-0.170**	-0.155**	-0.128**	-0.129**	-0.123**
	(0.0111)	(0.0110)	(0.0115)	(0.0120)	(0.0115)
Shock year+6	-0.166**	-0.150**	-0.120**	-0.123**	-0.115**
	(0.0126)	(0.0125)	(0.0129)	(0.0135)	(0.0129)
Shock year+7	-0.168**	-0.151**	-0.120**	-0.123**	-0.113**
	(0.0140)	(0.0139)	(0.0143)	(0.0151)	(0.0143)
Shock year+8	-0.158**	-0.142**	-0.108**	-0.112**	-0.101**
	(0.0158)	(0.0157)	(0.0161)	(0.0170)	(0.0161)
Shock year+9	-0.135**	-0.120**	-0.0853**	-0.0892**	-0.0771**
	(0.0182)	(0.0181)	(0.0185)	(0.0195)	(0.0186)
Shock year+10	-0.125**	-0.111**	-0.0745**	-0.0644**	-0.0643**
	(0.0208)	(0.0207)	(0.0211)	(0.0236)	(0.0212)
# observations	5,545,502	5,545,502	5,545,502	5,545,502	5,545,502

Note: FE estimates using the pre-matched sample described in the data section. Outcome is log yearly labor earnings. Column 2 includes a dummy for a second shock, column 3 in addition an interaction between this dummy and level of education. In column 4 the effect of any second shock is allowed to vary by time since the second shock. Column 5 includes controls for any third shock. The models also include time fixed effects and separate linear trends for individuals with and without university education. Robust standard errors in parentheses. * and ** indicate significance at 5 and 1 percent levels.

6 Conclusions

Our paper adds to the literature on the origins of the socio-economic gradient in health by examining if there is heterogeneity in the effects of health shocks across groups of different socio-economic status. Our results suggest that individuals with low education suffer disproportionally from a given health shock and that this pattern becomes more pronounced as people age. We are among the first to establish that there exists substantial heterogeneity by education in the labor market response to health shocks. The results are in line with the recent ones of García-Gómez et al. (2013), who found substantial heterogeneity by income in the effect of health shocks using Dutch data.

We also show that an important reason for the heterogenous income effects is that low-educated individuals are more likely to take up disability insurance, sickness insurance, unemployment insurance after experiencing a health shock, suggesting that their working situation makes them less able to cope with adverse health events. This is partly due to different sectors of employment, but we found no evidence that it was driven by heterogeneity in moral hazard, differences in treatment quality, and/or differential survival across groups. It suggests that differences in adherence to medical treatment across socioeconomic groups is an important explanation and we leave this to future research.

The existence of large heterogeneity in the impact of health shocks on labor outcomes means that policy advice that is based on average estimates may be misguided. Our results show that there may be gains in considering heterogenous effects, for instance, when evaluating economic gains of new medical technologies and treatments. By considering heterogenous effects, it may be possible to identify subgroups where the treatments have beneficial cost-benefits ratios. Targeted interventions towards such groups may thus lead to a more efficient use of health care resources.

References

Abadie, A. & Imbens, G. W. (2006). 'Large sample properties of matching estimators for average treatment effects'. *Econometrica*, 74(1), pp. 235 - 267.

Adams, P., Hurd, M. D., McFadden, D., Merrill, A. & Ribeiro, T. (2003). 'Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status'. *Journal of Econometrics*. 112(1), 3 - 56.

Antonovsky, A. (1967). 'Social class, life expectancy and overall mortality', *The Milbank Memorial Fund Quarterly*. 45(2), 31 - 73.

Au, D. W. H., Crossley, T. F. & Schellhorn, M. (2005). 'The effect of health changes and long-term health on the work activity of older canadians'. *Health Economics*, 14(10), 999 - 1018.

Case, A. & Deaton, A. (2005a). 'Health and wealth among the poor: India and South Africa compared'. *American Economic Review*, 95(2), 229 - 233.

Case, A. & Deaton, A. S. (2005b). Broken down by work and sex: How our health declines. In David Wise (ed.), *Analyses in the Economics of Aging* (pp. 185 - 212). Chicago: University of Chicago Press.

Dano, A. M. (2005). 'Road injuries and long-run effects on income and employment'. *Health Economics*, 14(9), 955 - 970.

Deaton, A. S. & Paxson, C. (2001). Mortality, education, income and inequality among American cohorts. In D. A. Wise (ed.), *Themes in the Economics of Aging* (pp.129 - 165), Chicago: University of Chicago Press.

Disney, R., Emmerson, C. & WakeFIeld, M. (2006). 'Ill health and retirement in Britain: A panel data-based analysis'. *Journal of Health Economics*, 25(4), 621 - 649.

d'Uva, T. B., Doorslaer, E. V., Lindeboom, M. & O'Donnell, O. (2008). 'Does reporting heterogeneity bias the measurement of health disparities?'. *Health Economics*, 17(3), 351 - 375.

Frijters, P., Haisken-DeNew, J. P. & Shields, M. A. (2005). 'The causal effect of income on health: Evidence from German reunification'. *Journal of Health Economics*, 24(5), 997 - 1017.

García-Gómez, P., van Kippersluis H., O'Donnell, O. & van Doorslaer E. (2013). 'Longterm and spillover effects of health shocks on employment and income'. *Journal of Human Resources*, 48(4), 873 - 909.

García-Gómez, P. & Nicolás, A.L., (2006). 'Health shocks, employment and income in the Spanish labour market'. *Health Economics*, 15(9), 997 - 1009.

Goldman, D. P. & Smith, J. P. (2002). 'Can patient Self-Management help explain the SES health gradient?'. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16), 10929-10934.

Halla, M. & Zweimüller, M. (2013). 'The effect of health and earnings: Quasiexperimental evidence from cummuting accidents'. *Labour Economics* 24, 23-38.

Lynch, S. M. (2003). 'Cohort and life-course patterns in the relationship between education and health: A hierarchical approach'. *Demography*, 40(2), 309 - 331.

Lundborg, P., Nilsson, M., & Vikström, J. (2011). *Socioeconomic heterogeneity in the effect of health shocks on earnings: Evidence from population-wide data on Swedish workers* (No. 2011: 11). Working Paper, IFAU-Institute for Labour Market Policy Evaluation.

Rosvall, M., Chaix, B., Lynch, J., Lindstrom, M. & Merlo, J. (2008). 'The association between socioeconomic position, use of revascularization procedures and five-year survival after recovery from acute myocardial infarction'. *BMC Public Health*, 8(44).

Schrijvers, C. T. & Mackenbach, J. P. (1994). 'Cancer patient survival by socioeconomic status in seven countries: a review for six common cancer sites [corrected]'. *Journal of Epidemiology and Community Health*, 48(5), 441 - 446.

Smith, J. P. (1999). 'Healthy bodies and thick wallets: The dual relation between health and economic status'. *Journal of Economic Perspectives*, 13(2), 145 - 166.

Sullivan, D., & Von Wachter, T. (2009). 'Job displacement and mortality: An analysis using administrative data'. *The Quarterly Journal of Economics*, 124(3), 1265-1306.

van Kippersluis, H., Van Ourti, T., O'Donnell, O. & van Doorslaer, E. (2009). 'Health and income across the life cycle and generations in Europe'. *Journal of Health Economics*, 28(4), 818 - 830.

Willson, A. E., Shuey, K. M. & Elder, Jr., G. H. (2007). 'Cumulative advantage processes as mechanisms of inequality in life course health'. *American Journal of Sociology*, 112(6), 1886 - 1924.

Appendix A Heterogeneous Effects by Type of Health Shock

Table A:1: Estimates of heterogenous effects by level of education and type of health shock (A)

	(1) Infectious	(2) Cancer	(3) Mental	(4) Nerve	(5) Heart
Shock year	-0.0487**	-0.104**	-0.0354	-0.0416	-0.0900**
	(0.0148)	(0.0152)	(0.0203)	(0.0232)	(0.0118)
Shock year+1	-0.0678**	-0.150**	-0.126**	-0.0991**	-0.186**
	(0.0200)	(0.0277)	(0.0315)	(0.0343)	(0.0187)
Shock year+2	-0.108**	-0.172**	-0.187**	-0.129**	-0.185**
	(0.0237)	(0.0294)	(0.0349)	(0.0385)	(0.0224)
Shock year+3	-0.121**	-0.221**	-0.196**	-0.162**	-0.198**
	(0.0272)	(0.0312)	(0.0382)	(0.0455)	(0.0261)
Shock year+4	-0.101**	-0.226**	-0.215**	-0.182**	-0.193**
	(0.0316)	(0.0357)	(0.0404)	(0.0526)	(0.0299)
Shock year+5	-0.147**	-0.223**	-0.191**	-0.209**	-0.181**
	(0.0353)	(0.0405)	(0.0449)	(0.0596)	(0.0346)
Shock year+6	-0.163**	-0.235**	-0.171**	-0.205**	-0.183**
	(0.0410)	(0.0466)	(0.0492)	(0.0667)	(0.0385)
Shock year+7	-0.165**	-0.242**	-0.155**	-0.203**	-0.231**
	(0.0477)	(0.0526)	(0.0548)	(0.0723)	(0.0423)
Shock year+8	-0.144**	-0.221**	-0.208**	-0.125	-0.195**
	(0.0553)	(0.0615)	(0.0586)	(0.0907)	(0.0468)
Shock year+9	-0.0589	-0.198**	-0.137*	-0.231*	-0.188**
	(0.0660)	(0.0706)	(0.0668)	(0.100)	(0.0530)
Shock year+10	-0.0952	-0.215**	-0.151*	-0.215	-0.142*
	(0.0753)	(0.0798)	(0.0741)	(0.111)	(0.0595)
# observations	238,508	218,496	351,033	125,171	984,003

Note: FE estimates using the pre-matched sample described in the data section. Outcome is log yearly labor earnings. The models also include time fixed effects and separate linear trends for individuals with and without university education. Robust standard errors in parentheses. * and ** indicate significance at 5 and 1 percent levels.

Table A:2: Estimates of heterogenous effects by level of education and type of health shock (B)

	(1) Respiratory	(2) Digestive	(3) Musculo-skeletal	(4) Genitourinary	(5) External
Shock year	-0.0441**	-0.0516**	-0.134**	-0.0446**	-0.104**
	(0.0130)	(0.00831)	(0.0158)	(0.0104)	(0.00848)
Shock year+1	-0.125**	-0.105**	-0.172**	-0.0812**	-0.167**
	(0.0178)	(0.0111)	(0.0236)	(0.0146)	(0.0125)
Shock year+2	-0.149**	-0.132**	-0.199**	-0.0969**	-0.135**
	(0.0216)	(0.0135)	(0.0256)	(0.0184)	(0.0151)
Shock year+3	-0.174**	-0.146**	-0.215**	-0.112**	-0.158**
	(0.0248)	(0.0166)	(0.0279)	(0.0209)	(0.0168)
Shock year+4	-0.204**	-0.129**	-0.187**	-0.136**	-0.178**
	(0.0284)	(0.0195)	(0.0325)	(0.0241)	(0.0185)
Shock year+5	-0.233**	-0.116**	-0.202**	-0.160**	-0.162**
	(0.0329)	(0.0237)	(0.0372)	(0.0273)	(0.0218)
Shock year+6	-0.214**	-0.131**	-0.185**	-0.201**	-0.165**
	(0.0381)	(0.0273)	(0.0411)	(0.0307)	(0.0247)
Shock year+7	-0.211**	-0.113**	-0.188**	-0.219**	-0.158**
	(0.0440)	(0.0309)	(0.0451)	(0.0343)	(0.0276)
Shock year+8	-0.221**	-0.0929**	-0.188**	-0.220**	-0.120**
	(0.0509)	(0.0357)	(0.0477)	(0.0400)	(0.0320)
Shock year+9	-0.233**	-0.0819	-0.163**	-0.216**	-0.106**
	(0.0584)	(0.0425)	(0.0555)	(0.0466)	(0.0374)
Shock year+10	-0.172*	-0.0906	-0.164*	-0.241**	-0.0925*
	(0.0726)	(0.0478)	(0.0637)	(0.0516)	(0.0435)
# observations	3,59,506	982,067	406,402	467,716	1,411,156

Note: FE estimates using the pre-matched sample described in the data section. Outcome is log yearly labor earnings. The models also include time fixed effects and separate linear trends for individuals with and without university education. Robust standard errors in parentheses. * and ** indicate significance at 5 and 1 percent levels.

Essay 2

Economic Incentives and Long-term Sickness Absence: The Indirect Effect of Replacement Rates on Absence Behavior

1 Introduction

Sickness insurance (SI) and disability insurance (DI) represent a large share of public expenditure in the industrialized world. In 2008 the average cost of sickness and disability benefits among OECD countries was estimated to be 2 percent of GDP. Sickness benefits accounted for almost half of this cost (OECD, 2009). This is almost three times more than was spent on unemployment benefits (OECD, 2009). The cost of sickness benefits is mostly driven by long-term absence spells. For Sweden, 51 percent of ongoing sickness absence spells in December 2012 lasted more than 90 days (Försäkringskassan, 2013). In addition to the direct cost of long-term sickness absence, a majority of disability claimants enter the system after a long sickness absence spell (OECD, 2009).

There is considerable evidence showing that economic incentives matter for the take up rates of social insurance (see e.g. Meyer et al. 1995; Curington, 1994; Johansson & Palme, 1996, 2002, 2005; Henrekson & Persson, 2004; Ziebarth & Karlsson, 2013; Petterson-Lidbom & Thoursie, 2013). By now we know that high replacement rates increase take up rates of SI and the duration of short-term absence. However, less is known about the impact of replacement rates on long-term absenteeism. Two exceptions to this are Ziebarth (2013) and Hesselius & Persson (2007). In Ziebarth (2013), the author shows that a reduction of the replacement rate in Germany had no effect on long-term absenteeism. For Sweden, Hesselius & Persson (2007) show that an increase in the replacement rate for absence spells lasting longer than 90 days increased the duration of spells.

In order to add to this rather short list, I utilize a reform of the compulsory Swedish SI scheme that occurred at the beginning of 1991.

The 1991 reform reduced the replacement level in the SI scheme from 90 percent of foregone earnings to 65 percent for the 3 first days of a sickness absence spell and to 80 percent for day 4 - 90.[1] From day 91 onwards the rate remained at 90 percent. The reason for the unusual incentive structure was a desire to cut public spending without punishing severely sick individuals. The new insurance scheme only affected new absence spells. Individuals on ongoing sickness absence were able to remain in the old insurance scheme.

The reform had two separate effects on individual sickness absence behavior – a *direct* and an *indirect* effect.

The direct effect stems from changes in the direct cost of entering into the sickness absence scheme (work-to-absence transitions), or continuing an absence spell (absence-to-work transitions). For short absence spells, the *direct* cost - defined as the percentage share of earnings not replaced by the SI scheme

[1]For the remainder of the paper, unless otherwise stated, absence and absence behavior refer to sickness absence.

- increased from 10 to 35 percent. The direct cost of medium-length absence spells increased from 10 to 20 percent. For absence spells of more than 90 days, however, the direct cost remained at 10 percent. If economic incentives affect the sick-leave decision, work-to-absence transitions will decrease while absence-to-work transitions will increase for short and medium-length absence spells following the reform.

The indirect effect stems from changes to the indirect cost of returning to work. The indirect cost is caused by the risk of returning to work and having to start a new absence spell. The indirect cost is defined as the difference between the share of earnings not replaced by the SI scheme during an ongoing spell and the corresponding share of a new spell.

Prior to the reform, the SI scheme replaced 90 percent of previous earnings regardless of absence duration. This meant that the indirect cost was zero throughout the absence spell. However, as the replacement rate increased over time in the new SI scheme, so did the indirect cost of returning to work. For medium-length absence spells the indirect cost increased by 15 percent. The corresponding increase for long absence spells is 25 percent. If economic incentives play a role in the sick-leave decision, this implies that the indirect effect of the reform would lead to a reduced absence-to-work transition for medium and long-term absence spells.

As shown, the direct and indirect effect on absence-to-work transitions varies during the absence spell. For medium-length spells the effects even move in different directions. This shows how important it is to account for both these effects when evaluating the impact of a reform of this kind. If only the direct effects are included in the estimates, the results will be biased. Using detailed data on the complete account of all sickness absence spells during the period, I am able to estimate the *indirect* effect of the reform on sickness absence behavior. Emphasis will be on medium and long sickness absence spells.

The reform has been evaluated in Johansson & Palme (2002, 2005). Both papers use a small sample of blue collar workers. Johansson & Palme (2002) jointly model the impact of both the replacement level reform and a tax reform. They show that the increase in absence costs reduced work-to-absence transitions and prolonged absence durations. In a second paper, Johansson & Palme (2005) apply a DID approach, which compares absence spells that started shortly before the reform with absence spells that started shortly after the reform. For absence-to-work transitions, the authors find that the outflow from sickness absence increased for short absence spells (1 - 3 days), while it decreased for long absence spells (longer than 90 days). But for medium-length absence spells (4 - 90 days), the estimated effect is insignificant. However, Johansson & Palme (2005) are unable to separate the direct and indirect effect on medium absence spells. The insignificant estimate is thus the combined direct and indirect effect of the reform.[2] They also show that the work-to-absence transition rate decreased due to the reform.

In their analysis of the effects within the duration (i.e. the effects of the reform between day 4 and 89 from day 90 onwards) Johansson & Palme (2005) were not able to properly take into account the compositional changes in the population of sickness absentees after the reform. In other words, since the new SI scheme influences the individual's decision to begin a spell of sickness absence, the population on sick leave after the reform will have on average poorer health than the population on sick leave before the reform.

The main contribution of this paper is that I improve the identification strategy in Johansson & Palme, thereby allowing me to estimate the indirect impact of the reform. To this end, I use all Swedish employees during the period in question. To identify the impact I utilize the fact that the new insurance scheme only affected absence spells that started after the reform was implemented. By only looking at individuals that started an absence spell prior to the reform (i.e. individuals that were unaffected by the reform at the time they began their sickness absence spell) I am able to fully control for the compositional effects of the reform and to identify the indirect effect. The reason for this is that the direct effect of the reform only affects new absence spells, ongoing absence spells are unaffected. The indirect effect of the reform, on the other hand, affects absence spells initiated in the post-reform period as well as ongoing absence spells. Thus, by only using absence spells from the pre-reform period, I am able to identify the indirect effect on absence behavior.

Since I have data on the whole universe of Swedish employees during the period in question, I am also able to extend the work by Johansson & Palme. The richer data set allows me to obtain more precise estimates of the reform's indirect effect on medium and long absence spells and to perform a heterogeneity analysis of the differences in effects between white collar workers and blue collar workers.

I find that the indirect effect of the reform reduced absence-to-work transitions both among medium as well as long-term absentees. The results are in accordance with the theoretical predictions. For medium absence spells, the reform reduces the probability to exit an absence spell by on average 17 percent. The corresponding effect on long absence spells is a 10-percent decrease. In comparison with previous evaluations of the reform, the effect on long-term absence is in line with Johansson & Palme (2005). However, unlike Johansson & Palme, I am also able to show that the indirect effect of the reform on medium absence spells is substantial and even greater than for the longer spells.

I also find that the indirect effect on absence-to-work transitions is stronger among white collar workers than among blue collar workers. The difference is

[2] This is not an issue for the long absence spells. For these Johansson & Palme identify and estimate the indirect effect.

not surprising given that white collar workers were covered by supplementary sickness compensation through their collective bargaining agreement before the reform but not after. This meant that the cost of returning to work was amplified for white collar workers. Furthermore, I find that male claimants react more strongly to the reform than women. This is especially true for medium-term absenteeism. This result supports previous findings that men, on average, tend to react more strongly to economic incentives in an SI scheme (see e.g. Henrekson & Persson 2004; Johansson & Palme, 1996 ; Ziebarth & Karlsson, 2013). A claimant's sickness absence history seems to have little influence over the indirect effect.

The paper is structured as follows: Section 2 discusses the views of economists on health and sickness absence. It also reviews previous research. Section 3 constitutes a description of the aspects of the Swedish sickness insurance system relevant to this study, including the 1991 reform used in the analysis. Section 4 provides stylized theoretical predictions of the reform's effect on absence behavior. Section 5 describes the data and the sampling method used. It also explains the empirical strategy. Section 6 presents the main results. Section 7 presents the heterogeneous effects for absence-to-work transitions. Finally, Section 8 concludes the paper.

2 Health, Sickness Absence and Economic Incentives

Health and sickness absence are clearly closely related. Ill health is the reason for most sickness absence spells. However, not all health problems lead to sickness absence. Assume that an individual's health ranges from very poor to very good. If the health status is below a certain level it is impossible for any individual to fulfill a work task. Economic incentives have no effect on the absence behavior of this group. At the other end of the spectrum, if an individual's health is over a certain level, (s)he would not qualify as ill and no doctor would provide him/her with a certificate for sick leave.[3] Between these two clear-cut cases are a number of borderline cases where it is possible for an individual to attend work but it is also possible to report in sick. If the individual can choose between being on sickness absence or being at work and is indifferent between the two alternatives given their current health status, the SI replacement level, work tasks, etc., a change in the replacement level could, ceteris paribus, theoretically affect the decision whether to go to work or not (Holmlund, 1991).

Most of the previous economic research on sickness absence has therefore focused on the economic incentives of the sickness insurance scheme and other potential determinants of sickness absence such as cyclical fluctuations and em-

[3] This is based on the assumption that the physician is able to observe the individual's health status.

ployment protection legislation. From this we know that the unemployment rate and absenteeism are negatively correlated. This is due both to changes in the composition of the labor force as well as behavioral responses (see e.g. Arai & Thoursie, 2005; Askildsen et al., 2005). Another result is that workers increase their sick leave usage when full employment protection is provided (Lindbeck et al., 2006; Ichino & Riphahn, 2005; Riphahn, 2004). There is also considerable evidence suggesting that an increase in the generosity of sickness benefits tends to increase absence rates. For the United States, Meyer et al. (1995) show that workers' compensation for work-related injuries leading to temporary total disabilities led to an increase in absence duration in the 1980s. Curington (1994) finds mixed results on work absence behavior following a number of legislative changes on the benefit levels from 1964 - 1983. Johansson & Palme (1996) empirically reveal that the direct cost of being absent has a negative impact on work absence for a sample of male blue-collar workers. Henrekson & Persson (2004) uses aggregated Swedish data to study a number of legislative changes in the replacement rate to show that economic incentives have a strong impact on absence behavior. Petterson-Lidbom & Thoursie (2013) show that an increase in Swedish benefits levels in 1987 led to an increase in absence.

The above research is almost entirely empirical. Theoretical research on this topic is rarer. Some examples are Ehrenberg (1970), Barmby et al. (1994), and Coles & Treble (1996). Engström (2007) incorporates sickness absence in a general equilibrium model of search unemployment. They show that higher unemployment benefits increase absenteeism among employed workers.

There are fewer studies when it comes to long-term sickness absence and locking-in effects. In a study by Ziebarth (2013), the author assesses a reduction in the replacement rate in the German health insurance system. The results suggest that the reform had no effect on long-term absenteeism on the average population. Ziebarth does on the other hand find some heterogeneity in the effects and a small but significant decrease in absence durations for the poor and for middle-aged full-time employees. Hesselius & Persson (2007) evaluate a policy change to the Swedish SI scheme in 1998. After the reform, insurance claimants were allowed 10 percentage points additional compensation from collective agreements between day 91 and 360 of an absence spell. The results indicate that absence spells of at least 91 days increased by 4.7 days on average.

3 Swedish Sickness Insurance and the Reform

Sweden's sickness insurance system replaces income for workers who cannot perform their usual work tasks because of temporary illness. The SI scheme is financed through payroll taxes on wages and covers all employees whose employers pay payroll tax. The scheme is administered by the Swedish Social Insurance Agency (SIA).

The state's monitoring of the insurance is very lax, especially during the first seven days of an absence spell. During this period it is up to the worker to decide whether (s)he is ill and to what extent this warrants an absence from work. From the eighth day of the absence spell, a medical certificate from a physician is needed. A physician is only supposed to write a certificate for an individual who is in such poor health that (s)he is unable to fulfill his/her ordinary work tasks. However, even with a medical certificate, there is some evidence indicating that patients may exert a strong influence over their sickness absence since they are able to make their physician issue a certificate even though he/she would not normally recommend sickness absence for the patient in question (see e.g. Svärsudd & Englund, 2000; Englund, 2008). Based on the information on the medical certificate, the local Social Insurance Office makes a decision whether to sick-list an individual or not. In other words, it is the local Social Insurance Office that monitors the insurance and prevents abuse of the system.

The proportion of earnings paid to the worker by the SI scheme, and the employer's responsibility for sickness benefits has changed on several occasions during the last few decades. The reform I am examining was implemented on March 1st 1991.

This reform changed the replacement level for all insured workers. The scheme changed from a uniform compensation level of 90 percent of foregone earnings[4] from the first day of a sick spell onwards, to a scheme where the level depended on the claimant's absence duration. More precisely, after March 1st 1991 the compensation level of the SI was:

- 65 percent compensation level for day 1 - 3.
- 80 percent compensation level for day 4 - 90.
- 90 percent compensation level from day 91 onwards.[5]

The reform applied to new spells only, not retroactively to ongoing spells. For this group, replacement levels remained at 90 percent.

Besides changing the foregone earnings compensated through the SI scheme,

[4] The insurance only replaces earnings up to the social security ceiling of 7.5 price base amounts. The price base amount is a measure set by the Swedish Government a year at a time. The amount is calculated based on changes in the consumer price index. The price base amount has various uses, including ensuring that sickness benefits, study support, etc., do not decline in value because of an increase in the general price level. In 1991, about 7 percent of the labor force had labor earnings above the social security ceiling.

[5] The compensation was still limited by the social security ceiling of 7.5 price base amounts. In 1991 this amounted to 32,200 SEK, approximately €3,300.

the reform also addressed the claimant's total compensation level. Most workers on the Swedish labor market are covered by supplementary sickness compensation through their collective bargaining agreement.[6] In general, this additional insurance replaced about 10 percent of the foregone earnings (i.e. a significant number of workers had a complete coverage rate).[7] Following the 1991 reform, any extra compensation after the 90th day of sickness absence led to an equivalent reduction in SI benefits.

The significant changes to the SI scheme were one of several proposed budget cuts proposed by the Swedish government at the beginning of 1991. The suggested cuts were a response to the deep economic crisis in Sweden at the time. The unorthodox design of the new SI scheme was motivated by a desire to reduce public spending without economically punishing sick people.

The compensation levels stayed the same until April 1993 when the compensation rate for absence of more than 90 days decreased to 80 percent. At the same time a waiting period of one day was introduced into the SI scheme.

4 Theoretical Predictions

This section describes how the reform could theoretically have affected sickness absence behavior on two margins: work-to-absence transition and absence-to-work transition. The change in the cost of these transitions depends on whether the insured worker is absent and, if absent, when the absence spell started and the length of the absence spell. I follow Johansson & Palme (2005) and define the *direct* cost of being absent from work (i.e. exit a work spell or continue an absence spell) as the percentage share of earnings not replaced by the SI scheme. The *indirect cost*, the cost of *returning* to work, i.e. exiting an absence spell and risking having to begin a new absence spell, is defined as the difference between the share of earnings not replaced by the SI scheme during an ongoing spell and the corresponding share of a new spell after returning to work.

Prior to the reform, the SI scheme paid 90 percent of previous earnings regardless of the length of the absence. This meant that the indirect cost was zero throughout the absence spell. However, after the reform the indirect cost was zero only for short absence spells, if the duration was less than three days. For medium-length absence spells, durations between 4 - 90 days, the indirect cost increased to 15 percent (80 % - 65 %), and for long absence spells, longer than 90 days, the indirect cost was 25 percent.

The direct cost, on the other hand, was 10 percent throughout the absence

[6] The employee does not need to be a member of a union to be entitled to the supplementary compensation, as long as the employer has signed a collective bargaining agreement all employees are automatically covered.

[7] Since sickness insurance benefits were taxed differently compared to labor income this actually meant that people earned more while they stayed at home on sickness absence.

spell prior to the reform. After the reform, the *direct* cost of short absence spells increased from 10 to 35 percent, while the direct cost for medium absence spells went from 10 to 20 percent. For absence spells of more than 90 days the direct cost remained at 10 percent.

The changes in costs are summarized in Table 1.

Table 1: Summary of the changes in cost due to the reform

	Direct cost			*Indirect cost*		
Spell duration	Pre-reform	Post-reform	Diff	Pre-reform	Post-reform	Diff
			Ongoing spells			
1 - 3 days	10 %	10 %	0 %	0 %	25 %	25 %
4 - 90 days	10 %	10 %	0 %	0 %	25 %	25 %
91 - days	10 %	10 %	0 %	0 %	25 %	25 %
			New spells			
1 - 3 days	10 %	35 %	25 %	0 %	0 %	0 %
4 - 90 days	10 %	20 %	10 %	0 %	15 %	15 %
91 - days	10 %	10 %	0 %	0 %	25 %	25 %

Note:The direct cost is defined as the as the percentage share of earnings not replaced by the SI scheme. The indirect cost is defined as the difference between the share of earnings not replaced by the SI scheme during an ongoing spell and the corresponding share of a new spell after returning to work.

It is clear from the table that for work-to-absence transitions and for absence spells shorter than 4 days, only the direct effect should have affected absence behavior. The direct cost of exiting a work spell increased unambiguously following the reform. For absence spells shorter than 4 days, the *direct* cost increased from 10 to 35 percent. At the same time there was no change in the cost of returning to work. Absence-to-work transitions should have increased in this state. If employees reacted to economic incentives work-to-absence transitions should have decreased, and absence-to-work transitions should have increased between day 1 and 3 of an absence spell.

However, for medium and longer absence spells, the reform had a number of implications on absence-to-work transitions.

For individuals who started an absence spell prior to the reform (i.e. ongoing spells) the direct cost remained unchanged at 10 percent. The claimant kept receiving 90 percent of foregone earnings throughout the absence spell. Therefore, the reform should not have had any direct effect on absence-to-work transitions for this group. However, the cost of returning to work increased for ongoing absence spells. For all ongoing absence spells, regardless of length, the cost of returning to work changed from 0 to 25 percent. In other words, the indirect cost of absence increased as a result of the reform. If economic incentives affect the sick-leave decision, absence-to-work transitions should decrease.

For new absence spells, the implications were again ambiguous and depended on the length of the spell.

There was an increase in the direct cost for medium absence spells (4 - 90 days) from 10 to 20 percent. This should have led to an increase in absence-to-work transitions. However, there was also an increase in the indirect cost. From day 4 to 90 of an absence spell the cost of returning to work changed from 0 to 15 percent. If economic incentives affect absence behavior, this should lead to a decrease in absence-to-work transitions. Since the direct and indirect effects work in opposite directions, the reform effects on work-to-absence transitions are theoretically ambiguous for absence spells ranging between 4 and 90 days.

For absence spells of more than 90 days, the direct cost is unchanged. The cost of returning to work, on the other hand, increased by 25 percentage points. The relative cost of being absent for the long-term decreased as a result of the reform. Again, if individuals react to economic incentives, absence-to-work transitions should decrease for absence spells of more than 90 days. If the indirect effects were not taken into account, the theoretical prediction would instead be that the reform has no effect on long absence spells.

As shown, the direct effect and the indirect effect on absence-to-work transitions move in different directions, and vary during the absence spell. The a priori effect of the reform on absence-to-work transitions is thus unclear. However, by only looking at ongoing spells, I am able to isolate the indirect effect of the reform on sickness absence behavior.

5 Data and Empirical Strategy

The data used in this paper are created from population-wide registers in the IFAU-database. Specifically, I use data from the Sickness Benefit Register and the LOUISE database. The LOUISE database is administered by Statistics Sweden and covers the entire Swedish population aged 16 - 64. The database includes background variables such as age, gender, number of children in the household, marital status, area of occupation, and annual labor earnings.

The Sickness Benefit Register is administered by the Swedish Social Insurance Agency (SIA) and is an event data base with information on sickness insurance benefit payments for every individual who has been sick and entitled to sickness insurance benefits. The register records the start and end date for every sickness absence spell in Sweden between 1987 - 1991 where the individual was entitled to sickness benefits from the social insurance system. During the period in question, the SIA was responsible for benefit payments from the first day of absence onwards.[8] In addition to the start and end dates, the reg-

[8]As from January 1st 1992, responsibility for the SI scheme during the first 10 calendar days was taken over by the employer. Short-term sickness spells for employed workers are therefore no longer registered by the SIA.

ister includes information about daily benefit amounts, employment status at the beginning of spell, and whether the absence is full time or part time.

5.1 Empirical Strategy for Estimating the Indirect Effects

To investigate the indirect effect on absence behavior, I utilize the fact that the compensation level only changed for spells initiated after March 1st, 1991. The sample is thus constructed from the pre-reform period only. The sampling strategy allows me to handle the compositional effects created by the reform.[9]

In order to obtain the sample used in the estimations, I select a random sample of 5 percent of all new sickness absence spells in January and February 1991. I define the treatment group as individuals who started their absence spell during this period, i.e. just before the reform, but who exited it *after* the reform had been implemented, on March 1st, 1991. The comparison group, on the other hand, consists of individuals who both started and exited an absence spell in January or February 1991. The sample strategy allows me to compare absence-to-work transitions shortly before the reform with absence-to-work transitions shortly after the reform. Spells are censored after 365 days. I restrict the sample to individuals aged between 20 and 63. To account for seasonal effects, a corresponding sample of new sickness absence spells in January and February 1990 is also included in the analysis. The importance of seasonal effects is illustrated in Figure 1. The figure shows the day-by-day work-to-absence transition for the period 1990 - 1992. The dashed, horizontal line marks the reform date. Figure 1 clearly reveals a seasonal pattern in work absence. Not accounting for these patterns would distort the estimates. Including spells from 1990 in the analysis also allows me to identify the effect on long absence spells. The final sample consists of 195,000 absence spells.

It is also worth noting that Figure 1 shows that there is a distinct drop in the transition rate around the time of the reform. This drop did not occur around the same time in previous years. I see this as the first sign of the reform affecting absence behavior.

[9] Even though it is unlikely, there could still be a behavioral response to the reform using this sample. The reform was announced about two weeks prior to the implementation, potentially influencing people to start a sickness absence before March 1st 1991. I test this hypothesis by only using spells started before February 10th. The results are robust.

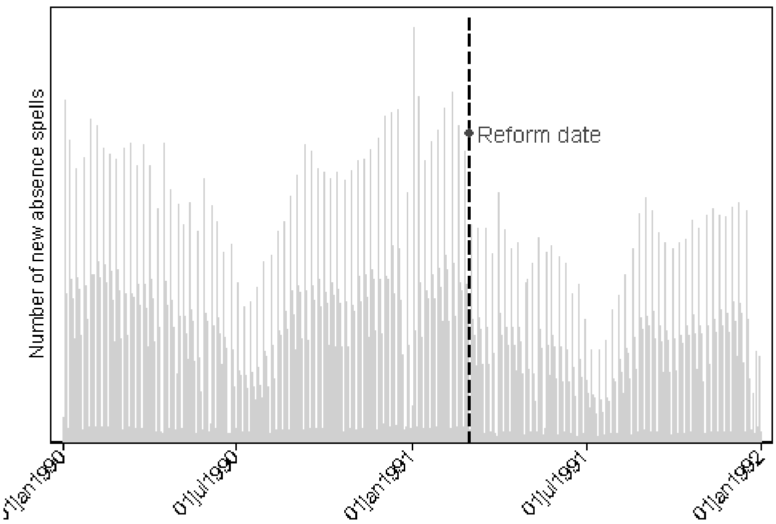


Figure 1: Day-by-day work-to-absence transitions.

In order to analyze the absence-to-work transitions, I estimate a piecewise constant Cox-proportional hazard (CPH) model. This strategy exploits the fact that the absence spell can be divided into three intervals, each having a different replacement rate, and thus entailing a different cost of returning to work. I let the impact of the reform vary over the two pre-specified time intervals: 4 - 90 days (medium) and more than 90 days (long).

To estimate the indirect effect on medium and long-term sickness absence, I use the following model specification:

$$\lambda_{(t)} = \lambda_0(t)exp(\gamma M + \beta_1 MD^R I(4-90) + \beta_2 MD^R I(91-)) \qquad (1)$$

where $\lambda_0(t)$ is the baseline hazard. M is a dummy variable taking the value one if the absence spell lasted beyond March 1st, 1990 or 1991 and zero otherwise. D^R is a dummy variable indicating whether the spell is treated, i.e. lasted beyond March 1991, or not. $I(j-j')$ are impulse functions, so that $I(j-j') = I(j \leq t \leq j')$ where I(.)=1 if the argument within parenthesis is true. The parameters of interest, β_2, and β_3 are the indirect effects for 4 - 90 and longer than 90-day durations respectively. If the reform had the impact outlined in Section 4 (i.e. absence-to-work transitions decreased as the indirect cost of absence increased) the β's should be less than one, indicating a decrease in the hazard rate and prolonged absence spells.

If there is an indirect effect on day 4 – 90 of an absence spell such that the outflow from absence is reduced, the estimated effect of β_2 will be attenuated towards one. The reason for this attenuation effect is simply that the comparison population remaining after day 90 has poorer health than the remaining treated population (affected by the economic incentives).[10]

An additional complication with the analysis is the recession that hit Sweden during the early 1990s. During the fall of 1991, the Swedish unemployment rate increased considerably. There is a literature that indicates that there is a relationship between the unemployment rate and the absence rate (see for instance Arai & Thoursie, 2005; Askildsen et al., 2005). I take this into account by using quarterly data from Statistics Sweden on the local labor market unemployment rates in January and February 1990 and 1991 respectively.[11]

In order to assess the robustness of the results, I will also perform a number of placebo regressions, i.e. move the reform one year back in time. These results are presented in Section 6.1.

6 Results

As stated in Section 4, the increase in the cost of returning to work could lower the absence-to-work transition rate, i.e. individuals on sickness absence would prolong their absence spell due to the risk of having to start a new absence spell and receive a lower compensation level. If this were the case, the reform would create a locking-in effect in the insurance.

Table 2 shows the estimates for the β_2 and β_3 from equation 1, capturing the indirect impact of the reform on absence-to-work transitions. I use four different model specifications in order to assess the robustness of the results. The first column presents the baseline model, controlling only for seasonal effects. The model specification in column 2 adds a vector of individual specific covariates and county fixed effects. Finally, in columns 3 and 4 I also control for the local labor market unemployment rate. The first row of the table reports the estimated impact on medium-length absence spells (4 - 90 days). The second row reports the same impact on long-term sickness absence spells (more than 90 days).

Looking at the table we see that the estimated coefficients are stable regardless of the model specification. Let us start with the impact on medium spells. In the baseline specification, column 1, the estimated coefficient is 0.833, a 16.7 percent decrease in the hazard. The estimated coefficient remains the same in column 2. When also controlling for socioeconomic status, the reform reduces the probability to exit an absence spell by, on average, 16.6 percent. The indirect effect increases slightly to around 17.1 percent when controls for local labor market unemployment rates are included in the model in columns 3 and

[10] This is known as a dynamic selection problem in the literature (see e.g. Van den Berg, 2001).

[11] I define the local labor market as the county of residence at the beginning of an absence spell. Sweden consisted of 24 counties during this period.

Table 2: Cox proportional hazard model estimates of the indirect effect on absence-to-work transitions

	(1)	(2)	(3)	(4)
Day 4 - 90	0.833**	0.834**	0.829**	0.829**
	(0.0140)	(0.0146)	(0.0147)	(0.0147)
Day 91 -	0.922*	0.904*	0.899*	0.898*
	(0.0381)	(0.0389)	(0.0388)	(0.0388)
County	No	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes
Unemployment	No	No	Yes	Yes
Unemployment^2	No	No	No	Yes
N	287,659	275,828	275,828	275,828

Note: Hazard ratios. The baseline hazard is specified as piecewise constant. Unemployment is measured at the county level. Controls include dummies for the level of education, age, sector of employment (2 digits), and gender. Absence spells are censored after 365 days. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$

4. All the estimated coefficients for medium-length absence spells are statistically significant at the one-percent level.

Let us now move on to the indirect effect on long absence spells which are spells of more than 90 days. Again, the estimated coefficients are fairly stable for all four model specifications. Using the baseline specification, the estimated coefficient is 0.922, a 7.78 percent reduction to the hazard. The second model specification returns an estimate of 0.904. For the most flexible model specifications - columns 3 and 4 - the estimated indirect effect again rises slightly (0.923 and 0.925). The indirect effect of the reform reduces the probability of exiting an absence spell of 91 days or longer by, on average, 10.1 percent when indicators related to the local labor market unemployment rate are included in the model. All estimated coefficients on the long absence spells are statistically significant at the five-percent level.

The main message from Table 2 is that the increase in the cost of returning to work caused a decrease in the absence-to-work transition. This is true for both medium and long absence spells. The results are coherent with the theoretical predictions of Section 4, when the cost of returning to work increases, absence-to-work transitions decreases. The indirect effects for both medium and long absence spells are estimated with great precision for both medium and long spells. All estimates are statistically significant at the five-percent level.

When comparing the size of the estimates, Table 2 shows that the indirect effect seems to be relatively smaller for the longest absence spell. The indirect effect on the probability of ending an absence spell after day 90 is weaker compared to the 4 - 90 day interval. A reason behind the weaker effect could be the dynamic selection described in Section 5.1. The reduced outflow between day 4 and 90 should attenuate the indirect effect on absence spells longer than 90 days towards one. The fact that, in spite of this attenuation bias, find a reduction in absence-to-work transitions for absence spells lasting longer than 90 days, strengthens the argument that indirect effects have an impact on absence behavior.

Additionally, economic incentives may have a more significant role for medium absence spells compared to longer ones. As health deficiencies lie behind most sickness absence spells, it is plausible to assume that the health of individuals with long absence spells is poorer than for those with shorter ones. Given this, the possibility to adapt behavior to a new insurance scheme might be more limited for this group.

6.1 Placebo Effects

In this section I provide placebo estimates for absence-to-work transitions. I do this by moving the reform one year back in time (i.e. before the actual reform took place). If such placebo effects emerge as significant, it would indicate that the previously estimated effects do not represent an effect of the actual reform. In the estimations I use a sample of 5 percent of all new sickness absence spells in January and February 1989 and January and February 1990. The treatment group now consists of individuals who exit their absence spells *after* March 1st, 1990.

Table 3 presents the placebo effects from the same four model specifications used in Section 6. By simply looking at the point estimates, the placebo regression suggests that there are small negative pre-treatment effects for medium absence spells, and small positive pre-treatment effects for long absence spells. All estimates are, however, insignificant at the five-percent level. The one exception is for medium absence spells in column 2, where the estimated coefficient is significant at the five-percent level.

Besides the fact that all estimates but one are insignificant, the estimates are also smaller compared to the main results. With this in mind, I see no reason to change my conclusions in Section 6.

7 Heterogeneous Effects

So far in the analysis I have assumed that the indirect effects had an impact on everyone in the same way. Clearly this is a restrictive assumption. In this section I therefore divide the sample by area of occupation (blue collar / white collar), gender, and by sickness absence history to see if there are heterogeneous reform effects in either of these dimensions.

Table 3: Cox proportional hazard model estimates of the placebo effect on absence-to-work transitions

	(1)	(2)	(3)	(4)
Day 4 - 90	0.980	0.963*	0.967	0.983
	(0.0169)	(0.0173)	(0.0173)	(0.0177)
Day 91 -	1.038	0.998	1.003	1.020
	(0.0431)	(0.0424)	(0.0426)	(0.0434)
County	No	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes
Unemployment	No	No	Yes	Yes
Unemploymentt^2	No	No	No	Yes
N	268,847	252,436	252,436	252,436

Note: Hazard ratios. The baseline hazard is specified as piecewise constant. Unemployment is measured at the county level. Controls include dummies for the level of education, age, and gender. Absence spells are censored after 365 days. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$

7.1 Heterogeneous Effects by Area of Occupation

Previous evaluations of the 1991 reform (Johansson & Palme, 2002, 2005) have focused solely on blue collar workers. The effect of the reform on white collar workers is thus unknown. This section, therefore, divides the indirect effect between blue *and* white collar workers. Blue collar workers are defined as individuals with less than three years of upper secondary education, while workers with three years of upper secondary education or more are defined as white collar workers.

I use the most flexible model specification, controlling for seasonal effects, a vector of individual covariates, the county and the local labor market unemployment rate, and run separate regression for white and blue collar workers respectively. The results are shown in Table 4, columns 1 and 2.

It is apparent from the table that the indirect effect of the reform was greater among white collar workers. These results are not surprising. During the period, white collar workers were covered by supplementary sickness insurance through their collective bargaining agreements. In general, this additional insurance paid out 10 percent of foregone earnings on top of the regular SI replacement rate. In other words, white collar workers had a complete SI cover rate prior to the reform. Since the reform also addressed the claimant's total compensation level, any extra compensation over 90 percent of foregone earnings led to an equivalent reduction in SI benefits, the cost of returning to work was thus higher for white collar workers.

Table 4: Cox proportional hazard model estimates of the effect on absence-to-work transitions - by area of occupation and gender

	(1) Blue collar	(2) White collar	(3) Men	(4) Women
Day 4 - 90	0.848**	0.787**	0.811**	0.860**
	(0.0175)	(0.0269)	(0.0193)	(0.0229)
Day 91 -	0.930	0.832*	0.912	0.883
	(0.0470)	(0.0693)	(0.0492)	(0.0640)
N	191,139	84,689	150,788	125,040

Note: Hazard ratios. The baseline hazard is specified as piecewise constant. The models include controls for the local labor market unemployment rate (first- and second order) and individual covariates (dummies for the level of education, age, sector of employment (2 digits), county of residence, and gender). Absence spells are censored after 365 days. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$

7.2 Heterogeneous Effects by Gender

It is a well known fact that there is a gender gap in sickness absence. Women are, on average, more absent from work for health reasons than men (see e.g. Paringer, 1983; Brostrom et al., 2002; Mastekaasa & Olsen, 1998; Angelov et al., 2013). However, men have been found to react more strongly to changes to the SI replacement levels (see e.g. Henrekson & Persson, 2004; Johansson & Palme, 1996; Ziebarth & Karlsson, 2013). In light of this, it is interesting to examine whether the reform effects differ between men and women.

To investigate whether there are heterogeneous responses according to gender, I run separate regressions for men and women. Columns 3 and 4 of Table 4 present the estimated indirect effects according to gender. I use the most flexible model specification, seasonal effects, a vector of individual covariates, the county, and the local labor market unemployment rate. It is clear from the table that, for medium length absence spells, the estimated indirect effect differs according to gender. Men reduced their probability to exit an absence spell by approximately 19 percent. The corresponding estimate for women is 14 percent. The estimates are significant at the one-percent level. For the longest absence spells, the point estimates indicate that men and women were affected to the same extent by the reform. However, neither of the estimates is statistically significant at the five-percent level.

To summarize, the results show that the indirect effect of the reform on absence-to-work transitions is greater for male workers compared to female workers. This difference is particularly evident in the medium length absence spells. The results are in line with previous research and lend support to the idea that men react more strongly to economic incentives and changes to the SI replacement levels compared to women.

7.3 Heterogeneous Effects by Sickness Absence History

Since the increased cost of returning to work hinges on the assumption of a return to sickness absence, the claimants' sickness absence history may have an impact on the reform effect. It is plausible to assume that individuals with a history of regular and/or long sickness absence spells have, on average, poorer health. Given this, these individuals will likely perceive a high risk of actually encountering the increase in the cost of returning to work (i.e. having another sickness absence). If this is the case, I should find a bigger reform effect among claimants with a history of sickness absence. Besides an increase in the risk of re-entering absence, individuals with a history of sickness absence may have acquired a greater understanding of how the SI system works and therefore have a better knowledge of what the cost of returning to work is. Again, this would suggest a greater reform effect for those who have utilized the SI scheme in the past.

I test whether sickness absence history has an impact on the indirect reform effect by running separate regressions by (1) number of absence spells, and (2) number of days on sickness absence in the two previous years. In both cases I divide the data into quartiles in order to make the groups equal in size. The estimates are presented in Panels A and B of Table 5. The estimates indicate that there is no clear relationship between sickness absence history and the indirect reform effect. This holds no matter what measure of sickness absence history is used.

Looking at Panel A of Table 5 - using the number of absence spells as a measure of absence history - the indirect reform effect on absence-to-work transitions seem to be strongest for individuals in the third quartile and weakest in the top quartile. The difference between quartiles one to three is, however, small. For long absence spells the relationship is even fuzzier. Only one of the estimates, the third quartile, is statistically significant. The lack of precision is not surprising given the smaller sample size used in the regressions. Looking at the point estimates, there is still no clear evidence of a relationship between absence history and the indirect effect among long-term absentees.

When I instead use the number of days as a measure of absence history - panel B of Table 5 - the indirect reform effect on medium-length absence spells is strongest among individuals in the bottom quartile (fewer than 13 days of absence) *and* in the top quartile (more than 66 days of absence). However, again, there is little difference between the quartiles. When it comes to long absence spells, the only statistically significant effect is for individuals with the most number of absence days. Looking at the point estimates, there is some evidence suggesting a stronger indirect effect among long-term absentees.

Based on these estimates, I find little support for the idea that the indirect reform effect is stronger among individuals who often utilized the SI scheme prior to the reform.

Table 5: Cox proportional hazard model estimates of the effect on absence-to-work transitions - by absence history

A. Number of absence spells	0 - 3	4 - 6	7 - 10	11 -
Day 4 - 90	0.829**	0.824**	0.812**	0.852**
	(0.0290)	(0.0303)	(0.0292)	(0.0296)
Day 90 -	0.911	0.929	0.828*	0.933
	(0.0771)	(0.0837)	(0.0742)	(0.0764)
N	71,381	66,460	66,567	71,420
B. Number of absence days	0 - 11	12 - 28	29 - 61	66 -
Day 4 - 90	0.837**	0.866**	0.863**	0.840**
	(0.0318)	(0.0326)	(0.0311)	(0.0266)
Day 91-	1.007	0.991	0.946	0.851**
	(0.110)	(0.118)	(0.0955)	(0.0509)
N	71,113	68,584	67,321	68,810

Note: Hazard ratios. The baseline hazard is specified as piecewise constant. The models include controls for the local labor market unemployment rate (first- and second order) and individual covariates (dummies for the level of education, age, sector of employment (2 digits), county of residence, and gender). Absence spells are censored after 365 days. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$

8 Concluding Remarks

The presence of moral hazards presents policy makers with a delicate problem: how do you balance the advantageous income-distribution properties of a generous sickness insurance system with the disadvantageous behavioral responses to such a system? In recent years, reductions in SI replacement levels have been widely used as a way of lowering absence rates. The cuts are justified by the assumption that economic incentives affect an individual's sick-leave decision. Increasing the cost of being absent would thus lower the absence rate.

In this paper, I have examined a reform of the Swedish SI scheme that took a somewhat different approach. Replacement levels were reduced, for all new absence spells, from 90 percent of foregone earnings to 65 percent during the 3 first days in an absence spell and to 80 percent for day 4 to 90 for all new absence spells. From day 91 onwards the rate remained at 90 percent. The reason behind maintaining the compensation level for spells of more than 90 days was to avoid negative economic effects on an already disadvantaged group.

I show that the reform had an indirect effect on absence behavior. The indirect effect (i.e. the cost of returning to work) increased due to the risk of having to start a new absence spell and receive the lower replacement rate. Using detailed data on the complete account of all sickness absence spells during the period, I am able to identify and estimate the indirect effect on absence-to-work transitions by comparing absence behavior before and after the reform.

The empirical findings confirm previous research about economic incentives and absence behavior; the absence-to-work transitions declined significantly after the reform. The indirect cost of returning to work seems to play an important role in the decision to end an absence spell. However, I find that the indirect effect was greater for medium-length absence spells compared to longer ones. Given that individuals on long-term sickness absence are severely ill, this could be interpreted as evidence of economic incentives playing a limited role when it comes to influencing long-term absence behavior.

I also find some heterogeneity in the results on medium and long-term absenteeism. White collar workers' reaction is stronger than that of blue collar workers. This is not surprising since the cost of returning to work was greater for white collar workers due to the fact that they also received supplementary sickness compensation through their collective bargaining agreement. Male claimants react more strongly to the reform than women. This is especially the case for long-term absenteeism. This supports previous findings that men, on average, tend to react stronger to economic incentives in the SI scheme (see e.g. Henrekson & Persson, 2004; Johansson & Palme, 1996 and Ziebarth & Karlsson, 2013). Additionally, I find that a claimant's sickness absence history seems to have little impact on the indirect effect.

References

Angelov, N., Johansson, P. & Lindahl, E. (2013). *Gender differences in sickness absence and the gender division of family responsibilities* (No. 2013: 9). Working Paper, IFAU-Institute for Evaluation of Labour Market and Education Policy.

Arai, M & Thoursie, P. (2005). 'Incentives And Selection In Cyclical Absenteeism'. *Labour Economics*, 12(2), 269-280.

Askildsen, J. E., Bratberg, E. & Nilsen, i. A. (2005).'Unemployment, labor force composition and sickness absence: A panel data study'. *Health Economics*, 14(11), 1087-1101.

Barmby, T., Sessions, J. G. & Treble, J. G. (1994). 'Absenteeism, efficiency wages and shirking'. *Scandinavian Journal of Economics*, 96(4), 561-566.

Broström, G., Palme, M. & Johansson, P. (2002). *Economic incentives and gender differences in work absence behavior*. Working Paper Series 2002:14, IFAU - Institute for Evaluation of Labour Market and Education Policy.

Coles, M. G. & Treble, J. G. (1996). 'Calculating the price of worker reliability'. *Labour Economics*, 3(2), 169 -188.

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

Ehrenberg, R. G. (1970). 'Absenteeism and the overtime decision', *American Economic Review*, 60(3), 352 - 357.

Englund, L. (2008). *Hur har distriktsläkares sjukskrivningspraxis förändrats under 11 år? Resultat av tre praxisundersökningar bland distriktsläkare i ett svenskt landsting*. Technical report, Centrum för klinisk forskning Dalarna, Sweden.

Engström, P. & Holmlund, B. (2007). 'Worker absenteeism in search equilibrium', *Scandinavian Journal of Economics*. 109(3), 439 - 467.

Försäkringskassan (2013). 'Sjuk- och rehabiliteringspenning - pågående sjukfall' , available on 2013-12-12, URL: http://www.forsakringskassan.se

Henrekson, M. & Persson, M. (2004). 'The effects on sick leave of changes in the sickness insurance system', *Journal of Labor Economics*, 22(1), 87 - 114.

Hesselius, P. & Persson, M. (2007). *Incentive and spill-over effects of supplementary sickness compensation* (No. 2007: 16). Working Paper, IFAU-Institute for Labour Market Policy Evaluation.

Holmlund, B. (1991). *Absenteeism , worker compensation, and the financing of health insurance*, Papers 1991s, Uppsala - Working Paper Series.

Ichino, A. & Riphahn, R. T. (2005). 'The effect of employment protection on worker effort: Absenteeism during and after probation'. *Journal of the European Economic Association*, 3(1), 120 - 143.

Johansson, P. & Palme, M. (1996). 'Do economic incentives affect work absence? Empirical evidence using Swedish micro data'. *Journal of Public Economics*, 59(2), 195 - 218.

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

Johansson, P. & Palme, M. (2005). 'Moral hazard and sickness insurance'. *Journal of Public Economics*, 89(9-10), 1879 - 1890.

Lindbeck, A., Palme, M. & Persson, M. (2006). *Job security and work absence: Evidence from a natural experiment* (No. 1687). CESifo working papers.

Mastekaasa, A. & Olsen, K. M. (1998). 'Gender, absenteeism, and job characteristics: A fixed effects approach'. *Work and Occupations*, 25(2), 195 - 228.

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

OECD (2009). 'Sickness, disability, and work'. *Background paper, High-level forum*, Stockholm. 14-15 May 2009.

Paringer, L. (1983). 'Women and absenteeism: Health or economics?'. *American Economic Review*, 73(2), 123 - 127.

Pettersson-Lidbom, P. & Thoursie, P. S. (2013). 'Temporary disability insurance and labor supply: Evidence from a natural experiment'. *Scandinavian Journal of Economics*, 115(2), 485 - 507.

Riphahn, R. T. (2004). 'Employment protection and effort among German employees', *Economics Letters*, 85(3), 353 - 357.

Svärdsudd, K., & Englund, L. (2000). 'Sick-listing habits among general practitioners in a Swedish county', *Scandinavian Journal of Primary Health Care*, 18(2), 81 - 86.

Van den Berg, G. J. (2001). 'Duration Models: Specification, identification, and multiple duration', *Handbook of Econometrics*, 5, 3381-3460.

Ziebarth, N. R. (2013). 'Long-term absenteeism and moral hazard—Evidence from a natural experiment'. *Labour Economics*, 24, 277-292.

Ziebarth, N. R. & Karlsson, M. (2013). 'The effects of expanding the generosity of the statutory sickness insurance system', *Journal of Applied Econometrics*.

Essay 3

Should Sickness Insurance and Health Care Be Administrated by the Same Jurisdiction? An Empirical Analysis

1 Introduction

Sweden has obligatory sickness and disability insurance which is both financed (from payroll taxes) and administered by the government. In order to receive sickness benefits from the Swedish Sickness Insurance Agency (SIA) the insured individual must have a certificate issued by a medical doctor (MD). For the government, it is important to have low and stable take-up rates as this ensures sustainable funding of the insurance. Health care is, however, administered at the county level, which means that monitoring is, to some extent, decentralized at a lower jurisdictional level than the funding and governance of the insurance. The advantage of this decentralization is the low cost to the government for monitoring. However, as the objective of health care providers is to allocate resources according to needs, rather than taking potential economic costs for the individuals or costs to the society into account, this may lead to the non-optimality (for government or society) of the amount of resources spent on monitoring the sick-listed individuals.

For the government it is important to have well-motivated certificates, as this reduces the monitoring cost for the SIA caseworkers. Writing well-motivated certificates may take time from the doctors' main tasks. Physicians have indicated that their role as experts in sickness certification is a frequently occurring and stressful task (Timpka et al., 1995). Since there is no special time devoted to the writing of certificates the motivation for spending time on the documentation and formulation of high quality certificates is often limited (see Alexandersson et al., 2009; Löfgren et al., 2007). If a caseworker does not have the necessary information in the certificate, he/she can re-remission the certificate back to the MD for completion. In addition to the extra cost for caseworkers and doctors, the re-remission process also incurs costs for the sick-listed individual as there are uncertainties about the payments of sickness benefits during the completion period. This could cause stress and potentially lead to longer sickness absences for the affected individual.

The purpose of this paper is not to perform a cost–benefit analysis of the decentralized monitoring system. It is rather more limited. The purpose is to analyze whether there are costs for the sick-listed individuals from having low quality certificates or, more precisely, of having the certificate re-remitted to the MD. If referring a certificate back prolongs the sickness absence there could be gains for both the government and for society to introduce incentives for doctors to write better certificates. To this end we use combined register and survey data. We find that re-remitting is correlated with the quality of the certificate, and moreover the length of the spell. Thus, this study also provides some evidence about the consequences of certificate quality. As far as we know, this is the first paper that addresses these issues.

There are, however, previous studies of the quality of medical certificates. A systematic review of the literature provided evidence that physicians frequently found it problematic to handle sickness certification tasks and that the certificates issued often is insufficient and of low quality (Wahlström & Alexanderson, 2004). More recent research has come to the same conclusion (see e.g. Hussey et al., 2004; Söderberg & Alexanderson, 2005; Einarsson, 2007; Löfgren et al., 2007). It has also been shown that doctors find it difficult to carry out their function as gatekeepers. Patients may claim sickness absence for very vague symptoms and the MD may have very limited knowledge about the patients' job, so the degree of reduced work capacity may be difficult to assess (see e.g. Svärsudd & Englund, 2000; Arrelöv, 2003; Englund, 2008). Furthermore, MD's feel that they lack competence in insurance medicine and have a limited knowledge of sickness and disability insurance regulations, and on how the certificates are being used by for example case workers) (Von Knorring, 2008; Gerner & Alexanderson, 2009).

We begin the paper by documenting that when we control for the diagnosis of the sick-listed individual there are no mean differences across socioeconomic status in re-remitting rates and quality. Next, we study the effect of requirement of completion on sickness absence. We find that re-remitting, on average, prolongs the sick-spell by approximately 30 percent. Our data allows us to perform suggestive tests of the reason for the observed effect. The results from these informal tests lead us to believe that the effect is due to decreased health caused by an increased stress related to uncertainty concerning current and future sickness benefit entitlement.

The yearly cost of low quality certificates in Uppsala County is approximately US$ 10 million (M) (72 M Swedish kronor (SEK)). The yearly cost of doctor visits related to sickness absence is around US$4 M. This suggests that giving doctors more in-patient time by, for instance, providing intergovernmental grants for high quality certificates would recoup these costs.

The remainder of this paper is structured as follows. Section 2 gives a brief summary of the Swedish sickness insurance system and the sick-listing process; Section 3 describes the data. This section also gives some descriptive statistics and explores the relationship between socioeconomic status and certificate quality; Section 4 gives the results together with a short discussion of the economic consequenses; Section 5 concludes.

2 The Swedish Sickness Insurance System

The sickness insurance system replaces income for individuals who cannot perform their usual work because of temporary illness. The level of sickness benefits and the employer's liability for sickness benefits has fluctuated in recent years. At the time of this study, employer's pay sickness benefits were equivalent to 80 percent of the worker's salary subject to a ceiling of $90 (655 SEK) per day during days 2–14 of the period of sick leave. After this period the responsibility for sickness benefits is transferred to the SIA.

During the first seven days of a sick leave, it is up to the individual to decide whether (s)he is ill and the extent to which this warrants absence from work. The individual merely has to inform the employer or the SIA that (s)he is sick. On the eighth day, a medical certificate is required. For sick leave that continues longer than two weeks, the employer notifies the SIA that the sick leave will continue. The SIA sends a letter to the insured with a form and a request for a medical certificate. A medical certificate is required for continued payment from the SIA (see Appendix A for an example of a medical certificate). The doctor indicates in the certificate the length and extent of sick leave needed. Based on the medical certificate, the SIA determines the right to sick leave–a process that normally takes one to two weeks after the end of the employer period. When this first sick leave period with benefits from the SIA has expired, a renewal certificate is issued if necessary. The renewal certificate is also sent to the SIA and a new assessment about the right to sickness benefits is conducted. If the renewal certificate expires and the insured is still sick, the process is repeated.

Based on the information in the medical certificate, the SIA decides whether the illness causes reduced work capacity. For those who have a job, the reduced work capacity is based primarily on their current job. For those who are unemployed the reduced work capacity should be assessed against jobs ordinarily available in the labor market. However, the proportion of cases in which the SIA decides against the doctor's recommendation is small. During 2006, the request for sickness benefits was rejected in 1.5 percent of all new cases. The percentage of rejections increased to 1.7 percent in 2008 (Försäkringskassan, 2007).

The assessment of entitlement is based on a guide (Försäkringskassan, 2004). The guide describes what information must be included in the medical certificate for the individual to be entitled to sickness benefits and to enable assesment of the need for rehabilitation. The SIA uses a support method in working with sick leave where a distinction is made between information that is "mandatory" and that which is "desirable" in the medical certificate. In situations where the case worker finds that the medical certificate does not contain sufficient information, they should re-remission the certificate back to the MD for completion.

Mandatory information is: the patient's name and social security number, the MD's name and clinic, the diagnosis or symptoms which are the basis for

[1]ICD-10, "International statistical classification of diseases and related health problems, tenth revision" is a coding of diseases and signs, symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases, as classified by the World Health Organization (WHO).

the reduced work capacity, and the diagnosis code according to ICD-10.[1] In addition there should be a description and medical assessment of the reduced work capacity. The doctor must also indicate findings from their examination in support of the diagnosis and the assessed requirement for vocational rehabilitation, if any. The medical certificate must also state whether the doctor's information is based on personal contact, telephone contact, journal entries, or other sources. The doctor should also give reasons why part-time sick leave (i.e., 25 percent, 50 percent or 75 percent reduced work capacity) and/or workplace rehabilitation is not possible. Finally, there should be a prognosis as to the insured's potential for regaining the capacity to work. Included under "desirable" information are such things as case history (i.e.) the insured's description of the illness and events that might have caused it. The Social Insurance Agency is not able to make a decision about eligibility if any of the *compulsory* information is missing and thus, the certificate should be referred back to the MD. In connection with the requirement of more information by the MD, the claimant is informed that his/her certificate has been returned to the MD and that no decision about sickness benefits can be made until the certificate is completed (Försäkringskassan, 2004).

A re-remission of the medical certificate means that the sick-listing process comes to a halt. Until a complete certificate reaches the case worker at the SIA, no decisions can be made about sickness benefit eligibility or potential rehabilitation. This may result in a locking-in effect for the claimant, thereby prolonging the absence spell. However, it is not fully apparent how important waiting time is for sick-absence. From Alexandersson et al. (2005) it may be concluded that physicians, as well as other occupational groups in the medical service, believe that waiting times do have consequences for patients. The waiting times are mostly associated with reappointment, treatment, and rehabilitation but also contact with case workers are considered problematic. Another negative aspect is that the uncertainty concerning entitlement may lead to the increased stress about potentially not receiving sickness benefits. This may affect the sick-listed individual's health negatively, thus prolonging the sickness absence.

3 Data

In this study we use data from two evaluations conducted by local social insurance office in the county of Uppsala. The aim of these evaluations was to examine the quality of medical certificates received by the office (see Appendix A for an example certificate). The reviewers examined all certificates received by the social insurance office in Uppsala during a two week period in 2006 and 2007. These certificates contain detailed information about diagnosis, recommended length of sickness absence, issuer, and, most importantly, information about whether the certificate was re-remitted for completion or not.

Certificates that contain all *compulsory* information are considered to be of *High Quality (HQ)* by the reviewers. No consideration was given to the information contained in such certificates.

The first evaluation took place during March 13–24 2006. During this period 786 certificates were collected and reviewed (Claesson, 2006). In the other evaluation, 1,127 certificates were examined during March 5–16 2007 (Einarsson, 2007). In total, we have information on 1,913 certificates, concerning 1,239 individuals. As certificates for prolonging a sickness absence spell are very different from new sick-listing, we removed these certificates from the analyses. After removal of the renewal certificates, 974 certificates remained. Out of these, 143 were re-remitted for completion.

We match (via a personal identification number) the information from the certificates with data from a set of administrative registers compiled by Statistics Sweden and the SIA. The data contains, besides a set of individual background characteristics, information about the total length of the sickness absence for the individual in connection with the studied certificate.[2]

3.1 Descriptive Statistics

Table 1 presents descriptive statistics for the socioeconomic and sickness-spell specific variables culled from the certificate. The descriptive statistics are presented separately for the group of individuals with re-remitted (column 2) and not re-remitted (column 1) certificates and for *HQ* (column 3) and not *HQ* (column 4) certificates.

From the top panel of Table 1 one can see that the only statistically significant differences in means between the individuals with re-remitted and not re-remitted certificates concerns *education unknown*.[3] From the bottom panel, where we have the sickness-spell specific variables, we find several statistically significant mean differences between the groups (see columns 1–2). For instance, the share of unemployed individuals is larger in the re-remission group. This difference comes as no surprise. It is, most likely, more difficult to assess the work capacity for unemployed individuals as their work capacity should be evaluated against the whole labour market.

Furthermore, re-remitted certificates are more often issued by occupational health care centers and less often issued at hospitals. One potential explanation for these mean differences is that patients in occupational health care centers

[2] Remember that the studied medical certificate is just the first certificate, hence the subscribed length in this certificate does not need to be the length of the sickness absence. The individuals are allowed to return early to work but can also stay on longer by using a prolonging certificate.

[3] It is worth pointing out that the level of education is unknown only for three individuals in the sample.

have illnesses with more vague diagnoses than patients from hospitals. In this case, the work capacity in relation to the diagnosis needs to be better documented. In the bottom panel, the diagnosis distribution is presented: the share of certificates with (i) mental and behavioral disorders *(behavioral)*, (ii) musculoskeletal system and connective tissue disorders *(musculoskeletal)*, (iii) other disorders *(other)* and (iv) *diagnosis missing*.

We can see that the share with a *behavioral* diagnosis is almost three times as large for the re-remission group than for the non re-remission group and that there are statistically significant smaller shares of *other* disorders in the re-remission group than in the non re-remission group. Since behavioral diagnoses often are vague, this result was expected.

Turning to the quality indicator (columns 3 and 4), in Table 1 we find unexceptionable many *HQ* certificates in the municipality of Tierp. Looking at the sickness-spell specific variables, we find a somewhat different pattern compared to the re-remission–non re-remission differences. The recommended sick leave is significantly shorter among the *HQ* certificates while there is no difference in employment status. *HQ* certificates are more often issued by occupational or primary health care physicians, and less often issued at a hospital. It is worth noting that for some of the *HQ* certificates, the issuer is unknown. The reason why these certificates still are judged as *HQ* could be that the reviewer was uncertain to what category the issuer belonged, and therefore reported it as unknown (Einarsson, 2007).

Looking at diagnoses, we see that the share of *behavioral* and *musculoskeletal* diagnoses are significantly higher among the *HQ* certificates. These differences could potentially stem from the fact that MDs know from experience that the caseworkers are stricter when it comes to these diagnoses. By being extra thorough when completing the certificate, MDs might hope to avoid a requirement of re-remission. It is slightly surprising that there is no statistically significant difference between the groups when it comes to *missing* diagnosis. A possible explanation of this is that the diagnosis may be indicated somewhere else on the certificate.

Table 1: Descriptive statistics of socioeconomic variables

	Re-remitted		HQ	
	No	Yes	No	Yes
Men	0.39	0.36	0.41	0.37
Age	45.5	43.8	45.3	45.2
	(11.8)	(12.0)	(12.1)	(11.7)
Married	0.49	0.50	0.50	0.48
Municipality				
Uppsala	0.51	0.44	0.50	0.50
Håbo	0.081	0.098	0.082	0.084
Älvkarleby	0.031	0.035	0.037	0.028
Tierp	0.073	0.063	0.045**	0.091**
Enköping	0.13	0.15	0.15	0.13
Östhammar	0.078	0.10	0.097	0.072
Heby	0.026	0.049	0.037	0.024
Knivsta	0.045	0.035	0.037	0.047
Children 0–3	0.12	0.13	0.11	0.13
Children 4–6	0.097	0.11	0.085	0.11
Children 7–10	0.097	0.12	0.11	0.096
Children 11–15	0.20	0.22	0.18	0.21
Children 16–17	0.088	0.070	0.065	0.100
Children in household	0.39	0.44	0.38	0.41
Primary School	0.19	0.18	0.20	0.18
High School	0.54	0.48	0.54	0.53
Upper Secondary	0.27	0.32	0.26	0.29
Education unknown	0.0012*	0.014*	0.0025	0.0035
Full time sick leave	0.87	0.85	0.89	0.86
Rec. sick-listing (days)	28.4	32.3	31.1*	27.5*
	(25.3)	(28.2)	(26.8)	(25.0)
Unemployed	0.060*	0.11*	0.052	0.079
Issuer				
Occupational physician	0.057*	0.10*	0.045*	0.077*
Hospital physician	0.37**	0.24**	0.42***	0.30***
Primary Healthcare physician	0.39	0.47	0.30***	0.47***
Private physician	0.084	0.077	0.065	0.096
Issuer unknown	0.097	0.10	0.17***	0.051***
Diagnosis type				
Behavioral	0.13***	0.34***	0.11***	0.19***
Musculoskeletal	0.25	0.24	0.22*	0.28*
Other diagnosis	0.61***	0.41***	0.65***	0.53***
Diagnosis missing	0.011	0.0070	0.017	0.0052
observations	974		974	

Note: Mean coefficients; standard errors in parentheses. *, ** and *** indicate significance at 10, 5 and 1 percent level, respectively.

3.2 Whose Certificate is Re-remitted?

A requirement for re-remission of a certificate is at the discretion of the case worker at the SIA. Because of this, there is a potential problem that the requirement of re-remission is based on individual characteristics because of statistical or preference discrimination.

In Table 2 we cross-tabulate *re-remitted* against *HQ,* where *re-remitted* takes the value 1 if the certificate is re-remitted for completion and 0 if not, and *HQ* takes the value 1 if the certificate is a *HQ* and 0 if not. The result from the cross tabulation is shown in Table 2. We see that the variables are not highly dependent.[4] The χ^2- test of independence is statistically significant at the 10 percent level only. As shown from this table, only about 59 percent of the certificates in the sample contains all compulsory information. In spite of this, only about 15 percent of the certificates are re-remitted. 7.7 percent of the *HQ* certificates are re-remittted in contrast with about 7 percent of the non *HQ*. This result is however not surprising since *HQ* is, basically, a minimum level of quality, often more information than this is needed for the entitlement decision. The reason for the 51 percent of certificates not re-remitted despite not being *HQ* is more difficult to explain. Einarsson (2007) suggests that the case worker could have received information from other sources than the certificate,[5] allowing the case worker to make a decision about entitlement.

Table 2: Distribution between level and treatment

		Re-remitted		
		No	Yes	Total
HQ	No	34.29	6.98	41.27
	Yes	51.03	7.70	58.73
	Total	85.32	14.68	100.00
chi2	2.727			
p	0.0987			

In order to study the more interesting conditional dependence (i.e., the dependence of the quality and re-remission when conditioning on the covariates) and to study if the decision of re-remission depends on socioeconomic factors (suggesting discrimination) we estimate three different logistic regression models.

The first model includes all socioeconomic and sickness-spell specific variables (see Table 1), except diagnosis and *HQ*. The second model adds the diag-

[4]We have also estimated a regression model where we regressed *HQ* on re-remission and R^2 is only 0.7 % in the regression.

[5]For instance from the claimant, their employer, or from the certificate, just not from the designated box.

noses and in the third model we also include *HQ*.

The first column of Table 3 presents the odds ratio from the first model specification. We can see that *age*, *education unknown*, *unemployed* and *recommended sick-listing* are statistically significant. Older individuals' certificates are less likely re-remitted while certificates for *unemployed*, those with *education unknown* and with long *recommended sick-listing* are more likely to be re-remitted for completion. However when we add the diagnosis (see column 2), all individual factors except *education unknown* are statistically insignificant and only certificates with a *behavioral* diagnosis are more often re-remitted. In other words, when conditioning on relevant information about the sickness, the claimant's socioeconomic background does not influence the caseworker's decision. The result that certificates with a behavioral diagnosis are more often re-remitted for completion than other certificates, all else being equal, is not very surprising. Behavioral diagnoses could be characterized as more diffuse and vague than other diagnoses. Previous studies have also found large variation in MD's sick-listing practices for these particular diagnoses. There is also a lack of knowledge of adequate treatments and rehabilitation for many behavioral diagnoses (Hensing & Wahlström, 2004; Alexandersson et al., 2005; Socialstyrelsen, 2003). From the third column of Table 3 we can see that when controlling for relevant socioeconomic variables as well as variables from the specific sickness absence, the *HQ* certificate has a significantly lower probability of being re-remitted.

To conclude, we find no support for the hypothesis that the individual's socioeconomic background affects the discretionary decision of re-remission by the caseworker. There is a strong conditional association between *HQ* and *re-remitted* which is why we believe *re-remitted* provides a (noisy) measure of the quality of the medical certificates.

Table 3: Linear probability model for the likelihood of re-remission

	(1) Re-remission	(2) Re-remission	(3) Re-remission
Age	0.981*	0.982	0.983
	(0.00917)	(0.00953)	(0.00959)
Men	0.934	0.923	0.918
	(0.189)	(0.189)	(0.189)
High School	0.831	0.827	0.845
	(0.216)	(0.220)	(0.227)
Upper Secondary	1.272	1.182	1.218
	(0.372)	(0.357)	(0.371)
Education unknown	12.65*	18.90*	20.92*
	(16.16)	(24.42)	(27.04)
Unemployed	2.002*	1.639	1.696
	(0.641)	(0.544)	(0.568)
Primary Healthcare physician	1.411	1.307	1.303
	(0.520)	(0.492)	(0.491)
Occupational physician	2.109	1.559	1.517
	(0.979)	(0.741)	(0.723)
Hospital physician	0.593	0.725	0.675
	(0.228)	(0.287)	(0.268)
Issuer unknown	0.958	1.023	0.837
	(0.425)	(0.463)	(0.386)
Full time sick leave	1.130	1.165	1.128
	(0.315)	(0.327)	(0.316)
Rec. sick-listing	1.021*	1.015	1.014
	(0.00896)	(0.00911)	(0.00921)
Behavioral		3.484***	3.735***
		(0.868)	(0.944)
Musculoskeletal		1.294	1.345
		(0.319)	(0.335)
Diagnosis missing		0.934	0.849
		(1.018)	(0.923)
HQ			0.601*
			(0.124)
Observations	973	973	973

Note: Odds ratio; Standard errors in parentheses. *, ** and *** indicate significance at 10, 5 and 1 percent level, respectively.

4 Re-remission and Sickness Absence

This section estimates the impact of re-remission for on sickness absence. Figure 1 shows the estimated Kaplan - Meier survival functions for the samples of re-remitted and non re-remitted certificates, respectively.

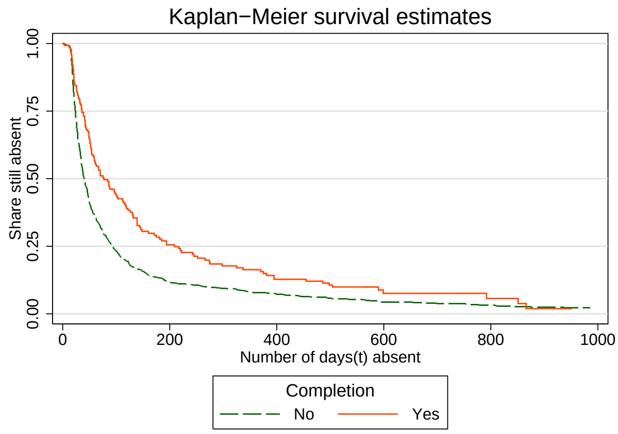


Figure 1: Fraction still absent due to sickness.

We can see that from day 20 of the sickness absence spell, the survival function for the individuals with re-remitted certificates is above that for individuals with certificates that are not re-remitted. This implies that the duration of sickness absence is, on average, longer for individuals with a re-remitted certificate compared to those whose certificates are not re-remitted. If treatment is the only thing differentiating the two groups, the difference between the survival functions could be interpreted as the effect of re-remission on the sickness absence duration. However, we have learned from Section 4 that there are observable differences between the two groups. These differences may also affect the sickness absence. In the following sub-section we will control for these observed differences by estimating Cox proportional hazard models.

4.1 Estimation and Results

Table 4 show the estimated effect from three different model specifications. The estimations are presented as hazard ratios, i.e., the relative risk of ending a sick spell.

Table 4: Cox proportional hazard model estimates of the effect of re-remission on sickness absence.

	(1) No controls	(2) No controls†	(3) Controls†
Re-remission	0.682***	0.658***	0.716***
	(0.0638)	(0.0704)	(0.0805)
Observations	974	973	973

†Stratified on recommended sick-listing.

Note: Hazard ratios. Controls include gender, marital status, number of kids in different age groups, level of education, immigrant status, age, residence municipality, employment status, diagnosis group, certificate issuer, recommended sick-listing, and degree of sick leave (full time / part time). Standard errors in parentheses. * , ** and *** indicate significance at 10, 5 and 1 percent level, respectively.

In the first specification, the Cox regression model is estimated without any control variables. With this specification, re-remission reduces the probability of ending a sickness-absence by, on average, 30 percent.

A key assumption in the proportional hazard model is that the hazards of the two groups are proportional at all durations. We know from Table 1 that *recommended days of sick-listing* is, on average, longer among the re-remitted certificates than among the no re-remitted ones. Recommended days could be viewed as a proxy for the severity of the illness (i.e., the longer the recommended sick-listing, the worse the illness). Hence, we find it plausible that the length of the sick-listing recommendation influences the probability of ending the sick-spell on a given day, all else equal. This violates the proportionality assumption. In order to handle this potential problem we stratify on *recommended days of sick-listing*. That is to say that we allow for separate baseline hazard functions for each value of the variable. Using this within recommended sick-listing variation enables us to compare the duration of re-remitted and non re-remitted individuals. The result from this stratified anlysis is shown in column 2. From this column we can see that the estimate gets slightly smaller (i.e., the re-remission effect increases).

For the last model specification we also add the control variables displayed in Table 1 into the model. The estimates from the stratified partial maximum likelihood estimator, including all control variables, are presented in column 3 of Table 4.[6] Based on these results we conclude that re-remitting the certificate for completion reduces the probability of ending a sick-spell by approximately 28 percent.

[6]The key assumption of proportionality is tested by analyzing the Schoenfeld residuals by following the generalization by Grambsch & Therneau (1994). The resulting test shows that the proportional hazard assumption cannot be rejected.

4.2 What Drives the Results?

We think of two, non-exclusive, reasons why the re-remission has such a large effect on the duration of the sickness absence. The first possible reason is that any decisions about rehabilitation or workplace adjustments cannot be made until the SIA has determined the entitlement. Thus, the re-remission could have a locking-in effect. The second possible reason is that individuals' health may be affected when they are informed that their certificates are re-remitted for completion. This information may create an uncertainty about the payment of sickness benefits and this could affect their health, which in turn may affect the length of the sickness absence. Below, we suggest two informal tests of these hypotheses.

The locking-in effect is tested by estimation of separate survival functions and hazard regression models for *HQ* equal to one and zero, respectively. The idea is the following: in contrast to the high quality certificates, the low quality certificates lack some compulsory information. A re-remission of a high quality certificate is thus more likely to depend on things that can be difficult to assess (e.g., how the patient's condition restricts their work ability or why they need to be on full time sick leave) and, hence, takes more time for the MD to do. This type of re-remission should reduce the outflow more from the sickness absence if the effect is due to a locking-in effect.

The estimates from the proportional hazard model are presented in Table 5. We see that the estimate is positive when the controls are excluded from the model and negative when they are added. Neither of them, however, are statistically significant. Thus, we find no support for the hypothesis of a locking-in-effect.

Table 5: Estimates of the effect of HQ-certificates, conditional on re-remission, using Cox proportional hazard method.

	(1) No controls	(2) Controls
HQ	1.140	0.986
	(0.309)	(0.237)

Note: Hazard ratios. Stratified on recommended sick-listing. Controls include gender, marital status, number of kids in different age groups, level of education, immigrant status, age, residence municipality, employment status, diagnosis group, certificate issuer, recommended sick-listing, and degree of sick leave (full time / part time). Standard errors in parentheses. *, ** and *** indicate significance at 10, 5 and 1 percent level, respectively.

If the results are caused by stress due to the uncertainty about sickness benefit entitlement, we believe that the re-remission effect should be larger for those with stress related or behavioral disorders. We test this hypothesis by estimating separate proportional hazard models for each diagnosis group. The results are presented in Table 6. From this table we see that the effect is large and statistically significant for those with a behavioral diagnosis. The hazard rate from sickness absence decreases by about 57 percent on average for this type of diagnosis if the certificate is re-remitted. This effect is almost twice as large as in the main analysis. For those with an *other diagnosis* the re-remission effect is also negative and statistically significant, while the estimate is positive but insignificant for those with a musculoskeletal diagnosis. These results support, though not conclusively, the idea that the re-remission effect stems from a health effect.

Table 6: Estimates of the effect of re-remission using Cox proportional hazard method, divided by diagnosis.

	(1) Behavioral	(2) Musculoskeletal	(3) Other diagnosis
Re-remission	0.428^{*}	1.142	0.603^{**}
	(0.146)	(0.357)	(0.113)

Note: Hazard ratios. Stratified on recommended sick-listing. The regression include controls for gender, marital status, number of kids in different age groups, level of education, immigrant status, age, residence municipality, employment status, diagnosis group, certificate issuer, recommended sick-listing, and degree of sick leave (full time / part time). Standard errors in parentheses. *, ** and *** indicate significance at 10, 5 and 1 percent level, respectively.

4.3 Economic Cost of Re-remission

The yearly cost for Uppsala County, in terms of added sickness benefits, for re-remitting medical certificates back for completion is calculated to be $10M (72 M SEK). [7] This is almost 7 percent of the total sickness benefits paid out in Uppsala County.[8] The yearly cost of doctor visits related to sickness absence is around $3.8M (27 M SEK).[9]

The average consultation time with a MD, including administrative work, was estimated at 32 minutes (Socialstyrelsen, 2005). Assuming that by increasing the consultation time by 10 minutes, MDs should be able to properly fill in

[7] Re-remitted sick spells have an expected duration of 180 days. The expected duration for the non-completion group is 116 days, a difference of 64 days. Taking the year 2006 as reference, there are $17,325$ sick spells requiring a medical certificate. Fourteen percent (see Table 2) of these certificates are expected to be re-remitted for completion. This means that 2245 ($0.14 \times 17,325$) certificates are estimated to be sent back for completion. The average sickness benefit during the period was $64 (464 SEK).

[8] During 2006, Uppsala County paid out almost $148 M (1,070 M SEK) in cash sickness benefits (Försäkringskassan, 2013).

[9] According to the Swedish Medical Association about 15 percent of all visits to primary health care physicians are related to sickness insurance (Jansson & Johansson, 2003). In Uppsala County, the number of visit in 2006 was 303,331 (SKL, 2007). This means approximately 45,500 (0.15*303,331) visits to primary health care are related to sickness insurance. The average cost of a doctor's visit in 2006 was $84 (604 SEK) (Socialstyrelsen, 2005).

the certificates and thus reduce the need for re-remission by 50 percent.[10] For Uppsala County this would imply an increased yearly cost of $1.2M (8.6 M SEK). This is considerably lower than the estimated cost of $5M (36 M SEK) from prolonged sickness absence due to re-remitted certificates. This then implies that there is room to allocate more working time for doctors for writing clear and well motivated certificates.

Since time is restricted, it is possible that MDs are unable to increase the consultation time and still fulfill their obligation to the other patients. Instead, the county could hire more MDs. Using the Swedish Structure of Earnings Survey, the yearly cost of a full time MD is estimated to be $117,000 (842,000 SEK) in 2006.[11] Even though other costs, such as administrative staff and local cost, may occur there is a still opportunity to allocate more working time for writing certificates and hire more MDs.

A directed intergovernmental grant with the aim of increasing the quality of the certificates would most likely mitigate the spillover effect and reduce the total cost for sickness absence in society.

5 Conclusions

Sweden has an obligatory sickness and disability insurance which is both financed (from payroll taxes) and administered by the government. However, in order for a sick individual to receive sickness benefits, (s)he needs a medical certificate issued by a MD. Based on the information in the certificate, a caseworker at the SIA decides whether the illness cause a reduced work capacity or not. If the caseworker does not have the necessary information in the certificate, they can re-remit the certificate back to the doctor for completion.

The main purpose of this paper has been to estimate the effect on sickness absence duration of having the medical certificate re-remitted to the doctor. The main result is that if a certificate is re-remitted, the sickness absence duration increases by on average 28 percent. Why do we find these large effects? We informally test for two causes: (1) decisions about rehabilitation or adjustments of the work place cannot be made until the SIA has determined the claimant's eligibility. In this context, the re-remission could be considered as a locking-in effect; and (2) the uncertainty concerning eligibility affects the recipient's health, thereby prolonging the sickness absence spell.

The locking-in effect is tested by estimating separate hazard regression models for those certificates that contain all compulsory information and for those who do not. The idea is that it should be easier to complete a certificate that lacks some compulsory information than certificates that contain all the compulsory information. A re-remission of a certificate that contains all the com-

[10] We consider re-remission as the lower bound of certificate quality.
[11] A monthly wage of $6958 (50,800 SEK) plus taxes (SCB, 2013).

pulsory information is more likely to depend on things that can be more difficult to assess and, hence, take more time for the MD (e.g., how the patient's condition restricts their work ability, or their need to be on full time sick leave). We find no support for this hypothesis.

The health effects hypothesis is tested by estimation of separate models for those with behavioral disorders or not. The idea is the following: if uncertainty about entitlement to sickness benefits affects the health status and thereby prolongs the sickness absence, we believe that this effect would be largest for those with a stress related or behavioral diagnosis. We find that the conditional probability of ending a sick-spell is 57 percent lower if the certificate is re-remitted for an individual with a behavioral diagnosis. The re-remission effect is almost twice as big in comparison with the full sample. Taking the maintained assumption of a larger effects for those with behavioral disorder, we believe that there is a "stress-effect" associated with having the certificates re-remitted for completion.

As we have shown, re-remission of the MD's certificate reduces the probability of ending a sickness absence spell. This however does not imply that SIA's role as gatekeeper in the social insurance system should be removed. Without this control mechanism the moral hazard problem associated with the insurance scheme would increase dramatically. This would have a long run negative effect on both the incidence and the prevalence of sickness absences. Such a development would be very costly for the government.

Instead, we propose the creation of directed intergovernmental grants from the state to the counties allowing MDs to spend more time with sickness absence patients. This allows the health care system to incorporate the cost of sickness absences into their decision making.

Additionally, it may be beneficial to cease informing claimants that their certificates have been re-remitted to the physician for completion. Since most claimants are eligible for sickness benefits once their certificate is completed, the risk of erroneous payments should be small. However, in case the claimant is not eligible, even after certificate completion, the claimant should be liable for reimbursement of the money. This situation of uncertainty with respect to payment would reduce the moral hazard in the sickness insurance and would increase the outflow especially among individuals with the best health. The reason for the last effect is simply that the cost for waiting (i.e. not working) for the final decision is higher for those with good health than for those with bad health given that the two groups have the same time preferences (Parsons, 1991).

References

Alexanderson, K., Brommels, M., Ekenvall, L., Karlsryd, E., Löfgren, A., Sundberg, L., & Österberg, M. (2005). *Problem inom hälso-och sjukvården kring handläggning av patienters sjukskrivning*. Stockholm: Sektionen för personskadeprevention, Karolinska institutet.

Alexandersson, K., Arrelöv, B., Hagberg, J., Ekmer, A., Lindholm, C., Löfgren, A., Nilsson, G., Skåner, Y., Söderberg, E. & Södersten, E. (2009). 'Läkares arbete med sjukskrivning: En enkät till alla läkare i Sverige 2008. Utveckling sedan 2004 i Östergötland och Stockholm', Sektionen för försäkringsmedicin, Karolinska institutet, Stockholm.

Arrelöv, B. (2003). *Towards Understanding of Determinants of Physicians' Sick-listing Practice and their Interrelations: A Population-based Epidemiological Sudy* (Doctoral dissertation, Uppsala University, Department of Public Health and Caring Sciences).

Claesson, F. (2006). *Läkarintygsrapport 2006*. Försäkringskassan, Uppsala, dnr 011973-2007.

Einarsson, L. (2007). *Medicinska underlag i Uppsala län - en studie av två myndigheters hantering av sjukintyg*, Försäkringskassan, Uppsala.

Englund, L. (2008). *Hur har distriktsläkares sjukskrivningspraxis förändrats under 11 år? Resultat av tre praxisundersökningar bland distriktsläkare i ett svenskt landsting*. Technical report, Centrum för klinisk forskning Dalarna, Sweden.

Försäkringskassan (2004), *Sjukpenning och samordnad rehabilitering (Vägledning 2004:2 version 9)*

Försäkringskassan (2007). *Nej till sjukpenning. Vad hände sen?*, Försäkringskassan Analyserar 2007:1.

Försäkringskassan (2013). Utbetalat belopp efter län och kön, 2000, http://www.forsakringskassan.se/statistik/statistik_och_analys/sjuk/, available on 2013-12-12.

Gerner U., & Alexanderson K. 'Issuing sickness certificates: a difficult task for physicians: A qualitative analysis of written statements in a Swedish survey'. *Scand J Public Health*, 37(1), 57-63.

Grambsch, P. M., & Therneau, T. M. (1994). 'Proportional hazards tests and diagnostics based on weighted residuals'. *Biometrika*, 81(3), 515-526.

Hensing, G. & Wahlström, R. (2004), Sickness absence and psychiatric disorders. In K. Alexanderson & A. Norlund (eds.), *Sickness absence - causes, consequences, and physicians' certification practice: A systematic literature review by the Swedish Council on Technology Assessment in Health Care. Scand J Public Health*, 32(suppl 63), 152-180.

Hussey, S., Hoddinott, P., Wilson, P., Dowell, J., & Barbour, R. (2004). 'Sickness certification system in the United Kingdom: Qualitative study of views of general practitioners in Scotland'. *BMJ: British Medical Journal*, 328(7431), 88.

Jansson, K., & Johansson, P. (2003). *Sjukvården och de sjukskrivna: Kan pengarna användas bättre?*. Sveriges Läkarförbund, Stockholm.

Löfgren, A., Hagberg, J., Arrelöv, B., Ponzer, S., & Alexanderson, K. (2007). 'Frequency and nature of problems associated with sickness certification tasks: A cross-sectional questionnaire study of 5455 physicians'. *Scandinavian Journal of Primary Health Care*, 25(3), 178-185.

Parsons, D.O. (1991). 'Self-screening in targeted public transfer programs'. *Journal of Political Economy*, 99(4), 859-76.

SCB (2013), Uttag ur Statistikdatabasen (Lönestrukturstatistik, landstingskommunal sektor),
http://www.scb.se/sv_/Hitta-statistik/Statistikdatabasen/
available on 2013-12-12.

Sveriges Kommuner och Landsting (2007). *Landsting och regioner i diagram och siffror 2006: Jämförelsetal för verksamhet och ekonomi för hälso- och sjukvården*, Sveriges kommuner och landsting (SKL), Stockholm.

Socialstyrelsen (2003).*'Utmattningssyndrom - Stressrelaterad psykisk ohälsa*. Socialstyrelsen, Stockholm.

Socialstyrelsen (2005). *Patientrelaterad redovisning av verksamhet och kostnader (KPP) inom primärvården: Metod, genomförande och användning med exempel från Borlänge sjukhus*. Socialstyrelsen, 2005

Svärdsudd, K., & Englund, L. (2000). 'Sick-listing habits among general practitioners in a Swedish county', *Scandinavian Journal of Primary Health Care*, 18(2), 81 - 86.

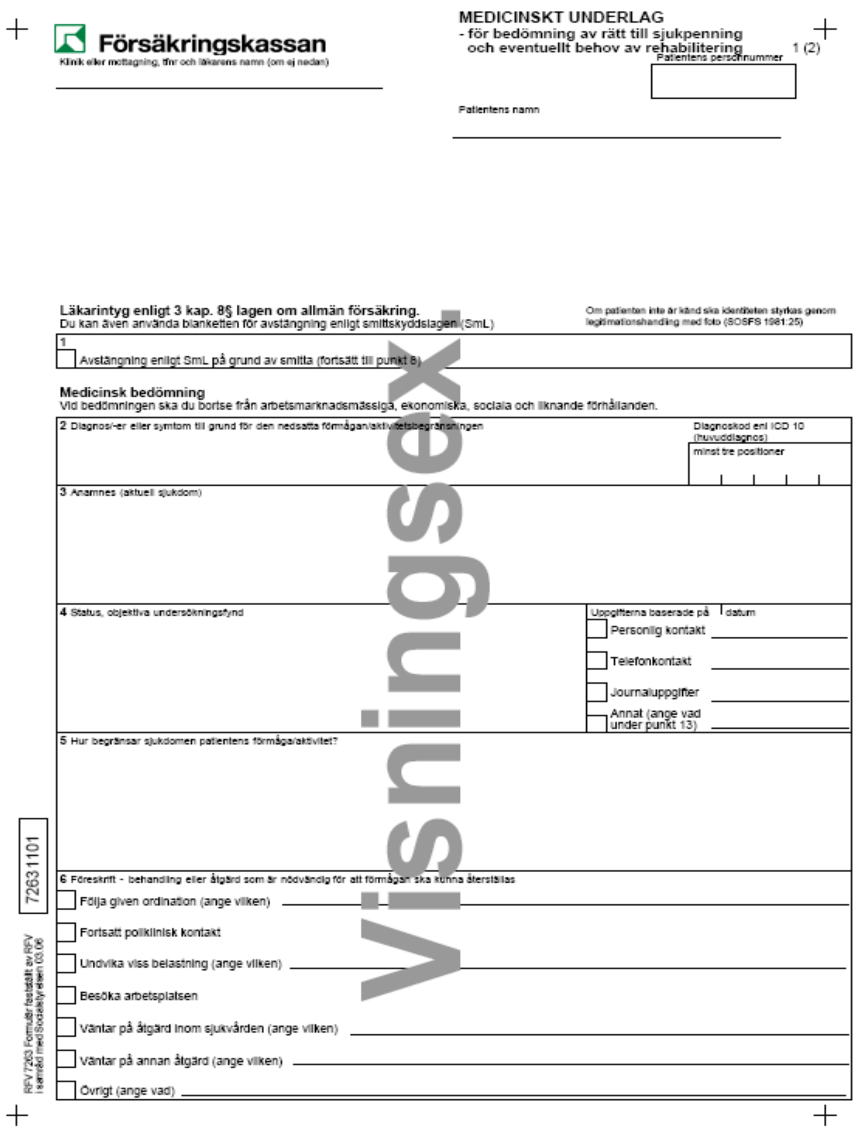
Söderberg, E., & Alexanderson, K. (2005).'Sickness certificates as a basis for decisions regarding entitlement to sickness insurance benefits'. *Scandinavian Journal of Public Health*, 33(4), 314-320.

Timpka, T., Hensing, G., & Alexanderson, K. (1995).'Dilemmas in sickness certification among Swedish physicians'. *The European Journal of Public Health*, 5(3), 215-219.

Von Knorring, M., Sundberg, L., Löfgren, A., & Alexanderson, K. (2008). 'Problems in sickness certification of patients: A qualitative study on views of 26 physicians in Sweden'. *Scandinavian Journal of Primary Health Care*, 26(1), 22-28.

Wahlström, R., & Alexanderson, K. (2004), Physicians' sick-listing practices. In K. Alexanderson & A. Norlund (eds.), *Sickness absence - causes, consequences, and physicians' certification practice: A systematic literature review by the Swedish Council on Technology Assessment in Health Care. Scand J Public Health*, 32(suppl 63), 222-255.

Appendix A Example Certificate



Försäkringskassan

Klinik eller mottagning, tfnr och läkarens namn (om ej nedan)

MEDICINSKT UNDERLAG
- för bedömning av rätt till sjukpenning och eventuellt behov av rehabilitering 1 (2)

Patientens personnummer

Patientens namn

Läkarintyg enligt 3 kap. 8§ lagen om allmän försäkring.
Du kan även använda blanketten för avstängning enligt smittskyddslagen (SmL)

Om patienten inte är känd ska identiteten styrkas genom legitimationshandling med foto (SOSFS 1981:25)

1
☐ Avstängning enligt SmL på grund av smitta (fortsätt till punkt 8)

Medicinsk bedömning
Vid bedömningen ska du bortse från arbetsmarknadsmässiga, ekonomiska, sociala och liknande förhållanden.

2 Diagnos/-er eller symtom till grund för den nedsatta förmågan/aktivitetsbegränsningen

Diagnoskod enl ICD 10 (huvuddiagnos)
minst tre positioner

3 Anamnes (aktuell sjukdom)

4 Status, objektiva undersökningsfynd

Uppgifterna baserade på | datum
☐ Personlig kontakt
☐ Telefonkontakt
☐ Journaluppgifter
☐ Annat (ange vad under punkt 13)

5 Hur begränsar sjukdomen patientens förmåga/aktivitet?

6 Föreskrift - behandling eller åtgärd som är nödvändig för att förmågan ska kunna återställas
☐ Följa given ordination (ange vilken)
☐ Fortsatt poliklinisk kontakt
☐ Undvika viss belastning (ange vilken)
☐ Besöka arbetsplatsen
☐ Väntar på åtgärd inom sjukvården (ange vilken)
☐ Väntar på annan åtgärd (ange vilken)
☐ Övrigt (ange vad)

72631101

RFV 7263 Formulär fastställt av RFV i samråd med Socialstyrelsen 03.06

Personnummer 2 (2)

7 Är arbetslivsinriktad rehabilitering aktuell?

☐ Ja ☐ Nej ☐ Kan inte bedömas för närvarande ☐ Behov av kontakt med företagshälsovård

8 Medicinsk bedömning av i vilken grad funktionsnedsättningen begränsar patientens förmåga

☐ att utföra sina vanliga arbetsuppgifter (ange arbetsuppgifternas art):

☐ om patienten är arbetslös; att söka/kunna utföra arbete som är normalt förekommande på arbetsmarknaden

☐ om patienten är föräldraledig med föräldrapenning; att vårda sitt barn

Arbetsförmågan bedöms

☐ delvis nedsatt med 1/4	fr.o.m. (år, mån, dag)	längst t.o.m. (år, mån, dag)
☐ delvis nedsatt med 1/2	fr.o.m. (år, mån, dag)	längst t.o.m. (år, mån, dag)
☐ delvis nedsatt med 3/4	fr.o.m. (år, mån, dag)	längst t.o.m. (år, mån, dag)
☐ helt nedsatt	fr.o.m. (år, mån, dag)	längst t.o.m. (år, mån, dag)

(om helt nedsatt, besvara frågorna nedan)

- Kan deltid vara olämplig av psykosociala skäl? ☐ Ja ☐ Nej
- Kan anpassade arbetsuppgifter möjliggöra sysselsättning på deltid/heltid? ☐ Ja ☐ Nej
- Kan deltid vara skadlig för sjukdomens förlopp? ☐ Ja ☐ Nej
- Kan deltid i nuvarande sysselsättning vara möjlig med hänsyn till symtom? ☐ Ja ☐ Nej
- Kan deltid förbättra prognosen för återgång i arbete? ☐ Ja ☐ Nej
- Kan deltidsarbete på annat sätt vara skadligt? ☐ Ja ☐ Nej

9 Prognos - bedöms patienten kunna få tillbaka sin förmåga till arbete/aktivitet?

☐ Ja, helt ☐ Ja, delvis ☐ Nej

10 Kan resor till och från arbetet med annat färdsätt än det patienten normalt använder göra det möjligt att återgå i arbete? ☐ Ja ☐ Nej

11 Önskar kontakt med Försäkringskassan ☐ Ja ☐ Nej

12 Önskar avstämningsmöte ☐ Ja ☐ Nej

13 Övriga upplysningar

Underskrift

14 Datum (år, mån, dag)

15 Läkarens namnteckning

16 Namn, mottagningens adress, telefonnummer (även riktnr) i klartext (om ej ovan)

Blanketten och mer information finns på www.forsakringskassan.se

72631201

RFV 7263 Formulär fastställt av RFV i samråd med Socialstyrelsen 03.06

Essay 4

Road Accidents, Lost Production and Social Insurance Expenditures: An Analysis Based on Administrative Registers

1 Introduction

Injuries due to road accidents represent a large welfare loss to society. According to the World Health Organization, road accidents account for 20 – 50 million injuries or disabilities, and almost 1.2 million deaths a year around the world (WHO, 2004). The high costs associated with road accidents have led to heavy regulations intended to create safer transport systems and reduce the number of accidents. To that end, governments invest in road infrastructure, law enforcement and traffic education. Understanding the cost of road accidents is critical for assessing the optimal use of these instruments. The cost of accidents, including both fatal and non-fatal damage costs, is usually defined as direct costs (for health care, property damage, and administration), indirect costs (for lost production and consumption), and a value for safety per se (immaterial costs, i.e., pain, grief, suffering, joy of life etc.). (For a standard methodological classification of road accident costs, see Alfaro et al., 1994, and Miller, 1997.)

The methods most frequently used to estimate these costs are the human capital approach (HCA) and the willingness-to-pay approach (WTP). In short, the HCA measures the loss to society due to an accident, based on the future production potential of the victim. The HCA is useful to estimate production losses, but should be viewed as an absolute lower bound of the costs of accidents. The WTP, on the other hand, estimates the value of life in terms of the amounts that individuals, or society, are prepared to pay to reduce risks to their lives. Although the HCA and the WTP are sometimes considered to be alternatives for valuing road traffic accidents, it should be noted that they estimate different cost components and that the two approaches are complementary (Wijnen et al., 2009). For society as a whole, the cost of road accidents is the total value of direct, indirect, and immaterial costs.

Since income varies and different methods of calculations are used, the estimated lost productive capacity due to a road accident differs between countries.[1] However, studies of the societal costs of road traffic accidents show that the productivity losses are a substantial part of the total costs. A share of 20 percent to more than 90 percent is no exception (see Elvik, 2000 and Trawén et al., 2002 for recent land comparisons). Such a big difference is due to variation in estimation methods regarding both the indirect and intangible costs.

Trawén et al. (2002) present at least four circumstances that influence the estimated lost production: (1) whether or not estimates exclude the value of the household market, (2) whether lost production refers to gross or net costs, (3) whether future loss of earnings are discounted and the level of the discount rate, and (4) the assumed level of growth in income. The estimated cost also

[1]For some recent, land specific examples see i.e. Bastida et al., 2004; De Brabander & Vereec, 2007; Finkelstein et. al, 2006; Connelly & Supangan, 2006.

depends on whether the estimate is incidence or prevalence based. Incidence based studies measure the cost of an illness from onset to end, whereas prevalence based studies measure the cost of an illness in one period, regardless of onset (Segel, 2006).

An additional factor that determines the monetary value of lost production is the road accident victim's productivity. The cost of lost output will ultimately depend on the share of victims in the work force, and their productivity. Most previous studies have relied on the average income of the population, only stratified by age and sex when determining the cost (see e.g. Finkelstein et al. 2006). The assumption that road accidents are random in the population is strong. However, there is some suggestive evidence that traffic injury is associated with socioeconomic status. Those who have low socioeconomic status sustain traffic injuries more often than those with high socioeconomic status (Graham et al., 2005; Hasselberg et al., 2005; Braver, 2003; Laflamme & Engström, 2002). If this is not taken into account, the indirect cost will be overestimated.

The purpose of this paper is to estimate the indirect cost (i.e. the value of lost production, of non-fatal road accidents in Sweden). In order to measure the production loss, I utilize the fact that all Swedish workers are covered by public sickness and disability insurance and measure the production loss as changes in labor market absence following a road traffic accident (see Section 3 for a description of the Swedish social insurance system). The costs are estimated with an incidence approach, which measures the cost of a road accident from start to end. The total cost is calculated as the present value of future earnings. This means that all costs associated with the road accident are assigned to the accident year and future costs are discounted.

For Sweden, Cedervall & Persson (1988) estimated the indirect cost of road accidents in 1985. The indirect cost of temporary absence is based on aggregated data from Statistics Sweden, while the cost of permanent absence is based on information from the Swedish Road Traffic Injuries Commission.[2] The monetary value of the lost production is based on Statistics Sweden's income survey for 1985. While there has been a lot of work done to add a valid value of safety per se to the cost estimates (see i.e. Persson & Cedervall, 1991; Persson et al., 2001; Andersson, 2005; Hultkrantz et al., 2006; Svensson, 2009), very little has been done on the indirect costs. The Swedish Transport Administration (STA)[3] uses an updated version of Cedervall & Persson (1988) (see i.e.

[2]The Swedish Road Traffic Injuries Commission is appointed by the Swedish Government. The commission works to ensure that road traffic accident victims will receive reasonable and uniform compensation for personal injury from insurance companies, especially in cases involving disability or death. The commission also issues remarks on remuneration in respect to personal injury to the courts or other authorities.

[3]The Swedish Transport Administration is responsible for road maintenance and road construction and for the execution of cost-effective road construction projects. Within this framework, prospective safety improvements are given explicit monetary values. These values are

Persson & Vegelius, 1995, Nilsson et. al., 1997) to calculate values for the indirect costs of road traffic accidents. In 2004, the net indirect cost of a serious, non-fatal, road traffic accident was valued at 261 000 SEK (€28,714)[4] by the STA (SIKA 2009).[5]

In order to improve on previous research, I use detailed Swedish longitudinal administrative data on sickness absence, hospitalizations, mortality, and socioeconomic status for the entire population of Swedish workers. The effect of road accidents on lost production is estimated within a difference-difference framework; I compare the change in labor market absence behavior over time between road accident victims and non-victims. The data and analytical approach gives me a number of advantages.

First, individuals involved in road accidents are identified through the Swedish National Patient Register (NPR). The register contains all accounts of in-patient treatment due to road accidents in Sweden during 2004. Using the NPR, I do not have to rely on police reported accident data, where there is substantial evidence of underreporting. This is especially true for non-fatal casualities (see e.g. Hauer & Hakkert, 1988; Elvik & Mysen, 1999; Rosman, 2001; Larsson, 2008). A problem with this identification strategy is that it does not identify individuals who do not seek medical care following a road accident. It will however give me a lower bound on the number of road traffic victims.

Second, my estimates of lost production are based on register data from the Swedish Sickness Benefit register and Statistic Sweden. These registers contain the start and end date for every sickness absence spell in Sweden, as well as annual labor earnings and disability insurance uptake. The data also contain detailed information on a number of individual specific background characteristics such as level of education, marital status, and labor market attachment. Using register data enables me to not only measure the number of work days lost due to a road traffic accident, but to also put a monetary value on lost production based on the victims' actual, pre-accident productivity. In other words, I do not have to rely on the assumption that the average productivity of road accident victims equals that of the population at large. The richness of the data also enables me to allow heterogeneity across age groups and between sexes.

Third, I am able to follow individuals for five years after the accident. Since the process of being granted disability benefits is long, and often following a

then considered, together with other cost and benefits (e.g. value of travelling time and noise), in the investment appraisal.

[4] 1 €= 9 SEK

[5] The STA account for the loss of consumption in their estimation of the Value of a Statistical Life. To avoid double counting, the consumption loss is deducted from the gross production loss which results in net production loss. The net production loss is assumed to make up 91 percent of the gross production loss of a serious road traffic injury. STA also adds VAT of 21 percent in order to transform factor prices to consumer prices. The charge of 21 percent corresponds to an average of different levels of VAT and other indirect taxes (SIKA, 2009).

longer period of sickness absence, this allows me to consider both the short and long-term impact of road accidents on temporary absence.

Finally, using detailed data on sickness absence and disability insurance uptake enables me to give an estimate of the burden road accidents have on the public sickness insurance system. Taken together, this allows me to estimate lost production not restricted to aggregated prevalence data and the tenuous assumption that accidents are random in the population.

This study finds that the risk of being involved in a road accident is far from random. Individuals with a history of unemployment, sickness absence, social insurance uptake, and lower incomes are more likely to be involved in these accidents. I find that involvement in a road traffic accident has a big effect on both temporary and permanent absence from work. The utilization of sickness insurance (SI) dramatically increases at the time of the traffic accident, the effect lingers on for three years. The impact on disability insurance (DI) take-up is also large: five years after the accident the probability of receiving DI benefits has increased by around 75 percent.

With this negative selection taken into account, the cost of a road accident – in terms of lost production – is approximately 900 million SEK (€90 million), or 142,000 SEK (€15,777) per accident. This is significantly lower than the value used by STA and what has been found in previous Swedish studies (MSB, 2009, Försäkringsförbundet, 2005). I see two main reasons for this. (1) I am not able to account for all production losses. Only sickness absences of more than 14 days are identified. The majority of absence spells are shorter than that, and if they were included in the cost calculations, estimated costs would increase. Furthermore, the cost of permanent disability among children and adolescents is not included in the estimates. (2) Previous studies have assumed that road traffic accidents are random in the population. This study shows that this is not the case. In fact, there is a very clear negative selection into road accidents. In order to test how important the assumption of road traffic accidents being random in the populations is, I re-calculate the value of lost production using the mean average income of the control group instead of the treated. This exercise show that assuming that road traffic accidents are random overestimates the value of lost production by approximately 19 percent or 175 million SEK (€19,4 million).

Finally, for the public insurance system, this study finds that the cost of road accidents is approximately 408 million SEK (€45 million).

The paper is structured as follows: Section 2 presents previous literature. Section 3 gives a short overview of social- and traffic insurance in Sweden. Section 4 describes the data and the sampling method used. This section also gives some descriptive statistics and explores the relationship between socioeconomic status and involvement in road traffic accidents. Section 5 explains the empirical strategy while Section 6 presents the estimated effects on sickness absence and disability insurance utilization. Section 7 and 8 presents the cost calculations. Finally, Section 9 concludes.

2 Previous Literature

The Swedish Civil Contingencies Agency (MSB, 2009) calculated all the costs that arose in Sweden in connection with road traffic accidents in 2005. The incidence cost of lost production due to non-fatal road traffic accidents among individuals aged 20 – 64 amounted to 2.3 billion SEK (€255 million). The corresponding prevalence cost was estimated at 4.1 billion SEK. The average net production loss, including VAT, was estimated at 342,946 SEK (€38,105) . The cost was estimated using a HCA. Future losses of earnings were discounted at four percent. An annual growth rate of one percent was assumed. The effect of road accidents on temporary absence was based on aggregated data from the Swedish Sickness Insurance Agency (SIA) and the Swedish National Board of Health and Welfare. The impact on permanent absence was based on aggregated data from the SIA and two prospective follow-ups by Maraste et al. (2002, 2003).[6] The monetary value of lost production is based on aggregated income data, stratified by age and gender, from Statistics Sweden.

In a report by Försäkringsförbundet[7] the prevalence cost of lost production in 2004, including payroll taxes of 40 percent, was estimated at 16 billion SEK. The estimate was based on the amount of lost income payments insurance companies made during 2004 to traffic injury victims (Försäkringsförbundet, 2005).

Beside the cost of road accidents, this paper also relates to the literature regarding the impact of health shocks on future labor market outcomes. Dano (2005) estimated the long run effects of road accidents on earnings and employment in Denmark. The overall result is that traffic injuries are associated with significant differences in labor market outcomes between injured persons and matched controls. Earnings are reduced in the long run for men and older women. Also, employment rates for injured men are significantly lower than for non-injured persons following the accident. For Sweden, Lundborg et al.

[6] Maraste et al., (2003) is a follow-up study of the healthcare consumption and health impairment of people injured in traffic. The study sample was made from people injured in traffic and registered for in-patient care during 1991/1992 at the Lidköping, Karlsham, Karlskrona, and Lund Hospital. In all, 230 people were followed for a period of 3.5-4 years after the accident. The authors show that, at the final subjective follow up, 23 percent of the patients still suffered some functional disability, pain and distress from the accident. Maraste et al., (2002) used the same sample of victims as in the study quoted above, but focused only on those with long-term subjective consequences three years after the accident. The study investigated the victims' healthcare consumption and lost productive capacity during an eight year period after the accident. The study found that 30 percent of the sample did not return completely to the labor market.

[7] Försäkringsförbundet is the industry organisation for insurance companies in Sweden. Their members account for approximately 90 percent of the insurance market.

(2011) estimated the impact of unexpected health shocks, including road traffic accidents, on future labor market outcomes in Sweden. The authors show that health shocks leads to an increased take up of disability, sickness, and unemployment insurance. They also find that health shocks incur long-lasting reductions to labor earnings. Halla & Zweimüller (2013) used a fixed-effects Difference-in-Differences approach to estimate the effect of accidents occurring on the way to and from work on future labor market outcomes. They found that after an initial period with a higher incidence of sick leave, injured workers are more likely to be unemployed or leave the labor market via disability retirement. Injured workers who manage to stay in employed incur persistent earnings losses.

3 Social Insurance and Traffic Insurance in Sweden

All employed and unemployed workers in Sweden are covered by public sickness (SI) and disability insurance (DI). Both insurances are administred by the Swedish Social Insurance Agency. Most workers are also covered by an unemployment insurance (UI) scheme. Unemployed individuals (either covered or not by the unemployment insurance scheme) have access to the sickness insurance scheme. Until July 2008, there was no formal time restriction on the length of sick leave in the sickness insurance scheme. However, such formal time restrictions existed in the unemployment insurance scheme.

Overall, the benefit requirements are the least generous in the unemployment insurance scheme and the most generous in the sickness insurance scheme. For example, the eligibility criteria in the unemployment insurance scheme imply that a significant share of unemployed in Sweden lack compensation from the unemployment insurance system.

3.1 Sickness Insurance

Sweden's sickness insurance system replaces income for workers who cannot perform their usual work tasks because of temporary illness. The SI is financed through payroll taxes on wages and covers all employees whose employers pays payroll tax. For the employed, the employer compensates absence for the first 14 days, with the first day being uncompensated. From day 14 and onward, the Swedish Sickness Insurance Agency (SIA) is responsible for benefit payments. For the unemployed, SIA becomes responsible on the second day of the sick spell. The benefit level in 2004 was 80 percent of foregone earnings, up to a cap

[8] The price base amount is a measure set by the Swedish government a year at a time. The amount is calculated based on changes in the consumer price index. The price base amount has various uses, including ensuring that sickness benefits, study support, etc., do not decline in value because of an increase in the general price level.

of 7.5 price base amounts.[8] In 2004 this amounted to 294,750 SEK (€32, 750). There was no time limit on sickness benefits.

3.2 Disability Insurance

Individuals 19 years and older are eligible for DI in Sweden. An individual is entitled to DI benefits if his/her working ability is assessed to be reduced by at least 25 percent for a period no shorter than one year. DI benefits can be either fully or partially granted (25, 50 or 75 percent) and they can be granted either permanently or on a time-restricted basis for up to 3 years. For individuals 29 years or younger, only time-restricted DI is granted. After a benefit period expires, the claimant's work capacity is reassessed. If it is still found to be reduced, a new period of DI is granted.

Full income-related DI amounts to 64 percent of an average of the claimant's previous gross yearly income. For those with low previous income, a fixed guarantee level applies. Full compensation at the guarantee level in 2004 corresponds to SEK 94,320 (€10, 480)

3.3 Unemployment Insurance

Unemployment benefits can be paid in two ways: a fixed basic compensation or an income-related amount based on previous earnings. To receive any compensation, the unemployed person must be at least 20 and fulfill: (1) the basic conditions, and (2) the work condition. The basic conditions are a set of rules for the unemployed. For instance, they state that he or she should be partially or completely unemployed and prepared to accept suitable job offers. The work condition specifies that the unemployed person must have been employed for approximately six of the past 12 months preceding unemployment. If these requirements are met, the unemployed person qualifies for fixed basic compensation. To be eligible for higher income-related compensation, a person also must have been a member of an unemployment insurance fund for at least 12 months preceding the first day of unemployment.

The replacement rate for the first 200 days is 80 percent of previous earnings up to a cap of a maximum of 680 SEK (€75) per day. Between day 201 and 300 (450), the compensation rate is 70 percent, and if the person participates in work training programs after day 300, the compensation is 65 percent of previous earnings.

3.4 Supplementary Benefits

In addition to compensation from the public transfer system, most employed people are entitled to supplemental compensation from agreements between social partners. For people on sick leave, these insurance policies top up the compensation below the cap up to a maximum of 90 percent of previous earnings. Above the cap, total compensation amounts to 80–90 percent of previous earnings for the first year of sick leave. About 90 percent of the working population is covered by these supplementary benefits (Sjögren-Lindquist & Wadensjö, 2006).

3.5 Traffic Insurance

All motor vehicles used on the road must be insured under the Motor Traffic Damage Act (1975:1410). The insurance is compulsory and provides compensation for third parties in a road traffic accident. It covers costs for anyone injured in an accident, including drivers and passengers. The insurance also replaces any lost earnings caused by the accident. However, compensation from the social insurance system is deducted from these benefits. This compensation may be reduced if the injured person has knowingly helped bring about the injury through intent, gross negligence or other negligence in conjunction with drunk driving. It also does not compensate for acts such as malicious damage, fire or theft.

4 Data and Sampling Strategy

This analysis is based on data from a set of administrative registers maintained by Statistics Sweden. In order to identify individuals involved in a road accident, I use the Swedish National Patient Register (NPR). The NPR includes information on all in-patient care in Sweden from 1987 to 2005. It includes information such as date of admission, length of stay, as well rich medical data including main and secondary diagnosis (through the International Classification of Diseases, ICD), and detailed information on medical procedures. Most importantly, the NPR contains information about external causes of injury or disease. Even though it is not possible to identify the individual accident through the NPR, using this information, I am able to identify those individuals who seek in-patient care due to a road accident. A problem with this identification strategy is that it does not identify individuals who do not seek medical care following a road accident.[9] However, this should give me a lower bound on the number of individuals involved in a road traffic accident.

I match this information with three other Swedish population wide registers; the LOUISE database, the National Causes of Death Register, and the Sick-

[9] Since the nominal fee for obtaining treatment in Sweden is very low I find it unlikely that a large number of individuals choose not to seek help in case of an road accident. This definition of a road accident victim is also in line with the definition used in the official statistics: A road accident refers to any accident involving at least one road vehicle, occurring on a road open to public circulation, and in which at least one person is injured or killed (Mattsson & Ungerbäck, 2013).

ness Benefit Register. The first of these, LOUISE, covers the entire Swedish population aged 16 - 65 and includes variables such as age, sex, immigration status, marital status, highest level of education, and yearly labor earnings for the years 2002 - 2010.[10] The National Causes of Death Register, records all deaths of individuals who have permanent residence in Sweden. The third register, the Sickness Benefit Register, is an event database with information on sickness insurance benefit payments for all people who have been sick and are entitled to sickness insurance benefits. The register records the start and end date for every sickness absence spell in Sweden between 1987 - 2010 that entitled sickness benefits from the social insurance system.

With this data I am able to use information on the outcomes of interest both before and after a road accident. Individuals who die within 30 days of admission are excluded from the sample. For individuals involved in more than one accident during the year, only the first accident is used in the analysis. Since the focus of this study is to investigate the impact road accidents have on lost labor production, I restrict the sample to individuals aged 20 - 59. The reason for this is that many of those younger than 20 have not yet entered the labor market, while those older than 59 are about to retire. By restricting the population to only those aged 20 - 59, I miss two large groups, the young and the elderly - approximately 45 percent of the total number of individuals seeking in-patient care due to a road accident. Children and adolescents (0 - 19 years old) represent around 28 percent of all admissions while those aged 60 or older represent 17 percent of cases (Larsson, 2008). The consequences of a road accident for these groups are obviously important, but the costs are, in many ways, different than those who are the focus of this study. These restrictions give me a sample of 7,005 road accident victims

The control group consists of individuals that potentially could have suffered a road accident in 2004 but did not. A random sample of 1 percent is used. The same age restrictions as the treatment group apply.

4.1 Descriptive Statistics

Previous research has found suggestive evidence that involvement in road traffic injuries is associated with lower socioeconomic status. Graham et al. (2005) show that the risk of pedestrian accidents increases substantially in deprived areas of the UK. For Sweden, Laflamme & Engström (2002) show that children whose parents were unskilled workers have a higher risk of being involved in a road traffic accident compared to children whose parents were employees with

[10] Labor earnings records all, gross, cash compensation paid by employers. Besides salary, this includes i.a. compensation paid by the employers during the first 14 days of a sickness spell. Sickness insurance benefits paid from the 15th sick day and onwards, unemployment insurance benefits, disability insurance benefits, and other forms of social benefits are not included in this measure.

intermediate or high salaries. Hasselberg et al. (2005) study the social background characteristics of drivers involved in accidents. They find that the incidence of road traffic accidents is higher in low-status groups than in high-status groups. Using U.S survey data, Braver (2003) shows that low socioeconomic status, indicated by level of education, was the strongest determinant of road accident fatalities. For Denmark, Christens (2001) has also found a negative relationship between income, education level and marital status and accident risk.

Table 1 provides descriptive statistics on the socioeconomic status for the sample used in this study. It is clear from the table that there are large differences between traffic accident victims in 2004 (treated) and those who were not accident victims in 2004 (controls).

Men are heavily overrepresented in the treatment group. Almost 67 percent of individuals involved in a traffic road accident were men. The treatment group is somewhat older than the comparisons on average (40.69 compared to 37.54). There is, however, a clear age pattern. The relative risk is the highest at the two youngest cohorts (ages 20-34). For older ages, the risk of being involved in an accident decreases with age. There are also large differences in the prevalence of road accidents by level of education. The fraction of individuals with primary education is significantly higher among the treatment group. Correspondingly, the percentage with an upper secondary or post graudate education is significantly lower. When examining marital status and children in household, it becomes clear that individuals with a child in the household and/or who are married are less likely to experience a road accident. Turning to labor market outcomes, we can see that road accident victims were, to a much lesser degree employed and were more frequently on sick leave and on DI benefits the year prior to the road accident. In 2003, 70.6 percent of the treated had employment and 77.1 percent for the control group, a difference of 16.5 percent. The numbers are similar in 2002. The average number of days on sick leave is 52 percent higher for road accident victims than for the controls. The share of DI recipients in 2003 are 36 percent higher among the treated than the controls.[11] Furthermore the share of welfare recipients is more than twice as high among the victims.[12] The victims also had lower labor market incomes on average prior to the traffic accident. This can be seen both in terms of average labor income and in the income distribution for 2003. The difference may be from the observed higher risk of unemployment, sickness absence and social assistance, but since the control group is more highly educated, the difference is also most likely due to lower wages.

[11] I only measure DI incidence, not how long the individual receives it or whether it is partially or fully granted DI.

[12] I only measure welfare incidence, not the amount of welfare received or how long the individual receives it.

Table 2 presents the results of a logit regression model in which the risk of being involved in a road accident is estimated. Taking other background characteristics into account, the risk of being involved in a road accident is twice as high for men than for women. The risk decreases with age. Being married and/or having children in the household also has a positive effect on the risk. Individuals with more education have a lower risk of being involved in a road accident compared to those with less education. Unemployment in the year prior to the accident has a negative association on the risk. Receiving welfare benefits in 2003 and/or 2002 increased the risk, as did being a DI recipient two years prior to the accident. Surprisingly, conditional on other background characteristics, neither labor income nor sickness absence are associated with a higher risk of being involved in a road accident.

To summarize, there is a clear negative selection in road accident involvement. Individuals in the treatment group are, on average, younger, less educated, and have a weaker connection to the labor market than the rest of the population. These results are in line with previous research and support the notion that traffic injury is associated with social status. The question is if these differences could be explained by differences in traffic exposure or just by differences in risk behavior.

Travel and communication behavior results from the Swedish National Travel Survey[13] are presented in Table A:1 in Appendix A. This table shows that individuals with a strong labor market attachment (i.e., higher education, employment and labor market income) are more exposed to traffic risk. In addition, individuals that are married or living in a household with children have more exposure. In other words, the differences found in Table 1 and Table 2 do not appear to be explained by differences in traffic exposure. From this analysis it seems plausible that differences stems from differences in risk behavior. Note that by risk behavior I also mean difference in modes of transport. Individuals with lower incomes may have financial restrictions which make them more likely to drive less safe cars than individuals with higher incomes. However, the budget restriction will also make them more likely to travel by public transport which would reduce the risk.

[13]The Swedish National Travel Survey, RES 2005-2006, contains information on the everyday movements and longer journeys made by Swedish residents. The survey was conducted on a daily basis during a one-year period, beginning in the autumn of 2005. In total, 27,000 interviews were conducted.

Table 1: Descriptive statistics for the control group and treatment group

	Controls	Treated	Diff
Share men	0.511	0.668	-0.157***
Mean age	40.69	37.54	3.144***
Age 20-24	0.104	0.194	-0.0894***
Age 25-34	0.232	0.254	-0.0221***
Age 35-44	0.262	0.232	0.0294***
Age 45-54	0.242	0.204	0.0373***
Age 55-60	0.160	0.115	0.0447***
Share married	0.406	0.286	0.120***
Children in household	0.408	0.340	0.0672***
Level of Education			
Primary education	0.161	0.216	-0.0547***
Lower secondary	0.505	0.556	-0.0510***
Upper secondary or post-grad.	0.322	0.218	0.104***
Edu. unknown	0.0116	0.00995	0.00166
Labor income			
Mean labor inc.	183,550	152,520	310,300***
Median labor inc.	185,300	141,800	435 000
1st quartile	0.237	0.297	-0.0596***
2nd quartile	0.248	0.271	-0.0235***
3rd quartile	0.255	0.233	0.0219***
4th quartile	0.260	0.199	0.0611***
Employment			
Employment rate, t-2	0.769	0.693	0.0767***
Employment rate, t-1	0.771	0.706	0.0654***
Sickness absence			
Average days on sick leave, t-2	12.07	17.49	-5.424***
Average days on sick leave, t-1	10.69	16.21	-5.519***
Welfare			
Welfare recipient, t-2	0.0425	0.0873	-0.0448***
Share welfare recipients, t-1	0.0536	0.111	-0.0577***
Disability Insurance			
Share DI recipients, t-2	0.0626	0.0850	-0.0224***
Share DI recipients, t-1	0.0715	0.0971	-0.0256***
N	48623	7005	-

Note: Education refers to the highest achieved level. Labor earnings records all, gross, cash compensation paid by employers. Days on sickness absence is net days with SI benefits. The individual is considered employed if he/she has a job in November of the given year. *, **, and *** indicate significance at 5, 1 and 0.1 percent level, respectively.

Table 2: Linear Probability model for the likelihood of being involved in a road accident.

	Odds Ratio	s.e
Men	1.924***	(0.0549)
Age	0.741***	(0.0345)
Age^2	1.007***	(0.00124)
Age^3	1.000***	(0.00001)
Married	0.838***	(0.0278)
Child in household	0.864***	(0.0266)
Education (reference category: Primary education)		
Lower Secondary	0.886***	(0.0310)
Upper Secondary or post-graduate	0.673***	(0.0286)
Education unknown	0.523***	(0.0705)
Labor market outcomes		
Labor income, t-2	1.000	(0.00003)
Labor income, t-1	1.000	(0.00003)
Unemployed, t-2	0.979	(0.0477)
Unemployed, t-1	0.897*	(0.0442)
Days on sickness absence, t-2	1.001**	(0.0003)
Days on sickness absence, t-1	1.002***	(0.0003)
Welfare recipient, t-2	1.316***	(0.0870)
Welfare recipient, t-1	1.538***	(0.0927)
Disability Insurance recipient, t-2	1.530**	(0.206)
Disability Insurance recipient, t-1	1.033	(0.133)
N	55,072	

Note: The estimation is conducted using a logit model. Education refers to the highest achieved level. Labor earnings records all, gross, cash compensation paid by employers. Days on sickness absence is net days with SI benefits. The individual is considered employed if he/she has a job in November each given year. Robust standard errors in parantheses. *, **, and *** indicate significance at 5, 1 and 0.1 percent level, respectively.

5 Empirical Strategy

In the econometric analysis the aim is to identify the effect a road accident has on lost production. Since the risk of being involved in an accident is far from being a random event, estimating the effects of accidents on social insurance benefits is difficult. However, since I have detailed data on the individuals before the potential accident, I can take into consideration the sorting or selection into an accident based on these observable characteristics. This selection on observables assumption is untestable and potentially strong. It may be the case that there are unobservables affecting both demand or requirements for social benefits and the risk of being in the accident. In this situation, any selection on observables estimator would most likely be biased. However, as I have observations on social benefits before the accident, I have the opportunity to control for these unobservables under the assumption of additive separability assumption (functional form) and a strict exogeneity assumption. That is, I specify a regression model in which I control for a specific accident group fixed effect. This group specific effect is assumed to be constant over the observation period. The strict exogeneity assumption means that conditional on this group effect and the covariates the accident is random, or in other word, the error term in the model is white noise.

To estimate the future take-up rate, I will use a Difference-in-Difference (DID) framework.

The "DID model" can be expressed as follows

$$y_{it} = h(\beta_0 + \beta_1 T_i + \tau_t + \delta D_{it} + \epsilon_{it}), \quad k = 2002, .., 2009, i = 1, ..., n. \quad (1)$$

Here T_i is one if an individual is a victim in 2004 and 0 otherwise. $h(.)$ is a functional form, D_{it} is a step function: it takes value one at the (t)ime period when the (i)ndividual is involved in the accident. For an individual not involved in an accident D_{it} is zero. Given the model assumption, δ measures the effect on being an accident victim on the outcome y. The critical assumption is that $\epsilon_{it}, t = 2002, 2003$ does not cause D_{it}.

Model 1 can easily be modified to include individual characteristics and to allow the effect to vary over time:

$$\begin{aligned} y_{it} = h(\beta_0 + \beta_1 T_i + \tau_t + \delta_k 1(t - j = k) D_{it} + \gamma' \mathbf{x}_i + \eta_{it}), \\ t = 2002, ..., 2009, i = 1, ..., n, k = k = 0, ..., 5. \quad (2) \end{aligned}$$

The coefficients δ_0,...,δ_5 capture the effect of a road accident in the accident year, one year after the year, and so on. $\gamma' \mathbf{x}_i$ is an index that characterizes observed indivdual heterogenity. The observed individual covariates are labor income, age, sex, children in household, family composition, and educational attainment. All covariates are measured one year prior to the treatment. This index removes potential time invariant differences in the outcome variable between individuals explained by observed individual characteristics. As $\epsilon_{it} = u_i + \eta_{it}$, where u_i is unobserved heterogeneity it may, conditionally on T_i, be positive correlated with D_{it}. The implication would be that the estimates of δ or δ_k are biased upwards if we do not add control variables to the regression model. The inclusion of the control variables can be used as a "test" of the DiD framework. The underlying idea is that if the estimates are plagued by selection bias (or omitted variables) the inclusion of covariates would moderate (or eliminate) this bias. If there are no differences in estimates with or without covariates then, given that unobserved variables are correlated with observed ones, unobserved heterogeneity should not plague the results. In practice, this idea has been around for quite some time: Altonji et al. (2005) provides a formal justification for such sensitivity analyses.

Since both outcomes of interest are non-negative, using a linear model is irrelevant (i.e. letting $h(\ ) = 1$). Instead, I chose a log-linear model when estimating the effect on days on sickness benefits and a logit model for the incidence (prevalence)[14] of disabilty benefits. I estimate the parameters in the following models

$$\ln E(y_{it}) = \beta_0 + \beta_1 T_i + \tau_t + \delta_k 1(t - j = k) D_{it} + \gamma' \mathbf{x}_i, \tag{3}$$

and

$$\ln[E(y_{it})/(1 - E(y_{it})] = \beta_0 + \beta_1 T_i + \tau_t + \delta_k 1(t - j = k) D_{it} + \gamma' \mathbf{x}_i \tag{4}$$

I estimate the parameters using pseudo maximum likelihood under Poisson and Bernoulli distributions, respectively. The estimator is consistent given the correct specification of the mean function. To handle problems with over dispersion and serial correlation, equation 3 is estimated with cluster-robust standard errors.[15]

One potential problem with the chosen strategy is that members of the control group are victims of road traffic accidents in the years after 2004. This would attenuate the estimates toward zero. As I have data on in-hospital care in 2005 I have the possibility to test for this potential problem. In 2005 2 percent of the control sample had an in-hospital care visit due to a road accident. This result provides a heuristic argument that the attenuation problem may be minor.

[14]We take into account prevalence of disability insurance in 2002 and 2003. An individual who is on disability insurance do not most often leave. This means that we estimate the effect on the yearly prevalence but the effect could be though of as incidence from 2003 to the year under interest.

[15]I estimate the model with the Huber/White/Sandwich estimator (see e.g White 1980).

6 Results

The results form the estimation of models 3 and 4 for days on sickness benefits and incidence (prevalence) to disability benefits. They are presented in Table 3, columns (1)-(2) and (3)-(4). The models without covariates are presented in column (1) and (3) respectively. The first message from the table is that estimates with and without covariates are basically indistinguishable with regards to sickness benefits. In terms of disability benefits, the effects are, unexpectedly, somewhat larger when adding control variables. The differences in estimates are, however, never statistically significant. Thus this provides an informal justification for the chosen framework.

Table 3: Estimates of the effect of road accidents on sickness absence, and disability insurance prevalence.

	Sickness absence		DI	
	(1)	(2)	(3)	(4)
2004	1.225***	1.230***	1.026	1.049
	(0.0442)	(0.0443)	(0.0217)	(0.0310)
2005	1.044***	1.047***	1.080**	1.149***
	(0.0536)	(0.0538)	(0.0295)	(0.0441)
2006	0.534***	0.537***	1.207***	1.395***
	(0.0568)	(0.0569)	(0.0382)	(0.0621)
2007	0.246***	0.249***	1.320***	1.643***
	(0.0609)	(0.0612)	(0.0453)	(0.0803)
2008	0.104	0.107	1.358***	1.758***
	(0.0654)	(0.0657)	(0.0486)	(0.0896)
2009	0.0446	0.0364	1.354***	1.750***
	(0.0707)	(0.0712)	(0.0495)	(0.0912)
Covariates	No	Yes	No	Yes
N	437,153	431,320	437,153	431,320

Note: The models include controls for age, marital status, children in household, level of education, labor earnings in 2003 and year fixed effects. Education refers to the highest achieved level. Labor earnings records all, gross, cash compensation paid by employers. Robust standard errors in parantheses. *, **, and *** indicate significance at 5, 1 and 0.1 percent level, respectively.

Starting with the effect on sickness absence we can see that traffic accidents have a very large direct impact. The number of days receiving sickness insurance benefits increases by, on average, 123 percent during the year of the accident. The drastic increase is not surprising given that the road accident victim is in need of inpatient care. This table also reveals that road accidents affect sickness absence in subsequent years. One year after the accident year, the number of days on sickness absence doubles due to the accident. The effect diminishes in year two and three, but sickness absence still increased by 50 percent and 25 percent, respectively through these years. There are no statistically significant differences in sickness absence four and five years after the traffic accident.

Examining the effect on disability insurance benefits in columns (3) and (4), we can see that there is no effect on DI uptake in the year of the accident. However, from 2005 (the year after the accident) and onwards, there is an increased inflow to DI. Given that the process of being granted disability benefits is long, and that replacement rates are higher in the SI system, this delayed effect is expected. The effect of road accidents on DI take-up is very large, five years after the accident the probability of receiving DI benefits increases, on average, by around 75 percent.

It is worth noting that people seem to transfer between the two insurance systems during the follow up period. The utilization of SI increases dramatically at the time of the traffic accident, an effect that lingers on for three years. However, as days with SI benefits decrease, DI benefits increase. Based on these estimates, it appears that individuals involved in road accidents are granted DI benefits following a longer period of sickness absence.

6.1 Heterogeneous Effects for Women and Men

A well known fact is that there is a gender gap in sickness absence. Women are, on average, more often absent from work for health reasons than men (see e.g. Paringer, 1983; Broström et al., 2002; Mastekaasa & Olsen, 1998; Angelov et al., 2013). It is therefore not unreasonable to believe that the effect of traffic accidents may vary by gender. I test this by running separate regressions for men and women.

The results from these estimations are presented in Table 4. From this table I can see that the immediate effect on sickness absence is stronger among men (see columns (1) and (2)). The number of days with sickness insurance benefits increases by, on average, 135 percent for men and 104 percent for women. The difference prevails one year after the accident. Examining the effect on inflow to DI benfits displayed in columns (3) and (4), I can see that men involved in a traffic accident are granted DI benefits at a faster pace than women. In 2005 the probability of receiving DI benifits increased by, on average, 10 percent for men while there is no statistically significant effect among women. However, from 2006 (two years after the accident) and onward the benfit take-up rate is slightly higher among women involved in a road accident. Three years after the accident, the probability of receiving DI benefits increased by 75 percent for women and 59 percent for men. However, there is no difference in DI take-up rate five years after the road accident.

Table 4: Estimates of heterogeneous effect of road accidents on sickness absence, and disability insurance prevalence

	Sickness absence		DI	
	Men	Women	Men	Women
2004	1.355***	1.042***	1.061	1.059
	(0.0614)	(0.0667)	(0.0385)	(0.0537)
2005	1.096***	0.954***	1.182***	1.125
	(0.0742)	(0.0814)	(0.0575)	(0.0708)
2006	0.516***	0.547***	1.387***	1.430***
	(0.0787)	(0.0852)	(0.0778)	(0.105)
2007	0.254**	0.250**	1.587***	1.754***
	(0.0847)	(0.0911)	(0.0977)	(0.140)
2008	0.106	0.0808	1.729***	1.818***
	(0.0894)	(0.100)	(0.111)	(0.151)
2009	0.0660	-0.0586	1.737***	1.777***
	(0.0958)	(0.110)	(0.115)	(0.149)
Covariates	Yes	Yes Yes	Yes	
N	228,741	202,579	228,741	202,579

Note: The models include controls for age, marital status, children in household, level of education, labor earnings in 2003 and year fixed effects. Education refers to the highest achieved level. Labor earnings records all, gross, cash compensation paid by employers. Robust standard errors in parantheses. *, **, and *** indicate significance at 5, 1 and 0.1 percent level, respectively.

7 The Indirect Cost of Road Accidents

The purpose of this section is to calculate the cost of lost production due to road accidents. The cost will be calculated using the human capital approach (HCA). The principle behind HCA is that of opportunity cost - indirect costs are equivalent to the production that would have been produced in the absence of the road accident. The total value of lost production depends on the amount of working life a person would reasonably expect to have and the worth of the labor to the workplace. Two key assumptions in the HCA are: (1) earnings reflect productivity - the economic value of an employee's productivity is equal to their labor earnings and payroll taxes; and (2) existence of a perfectly competitive labor market wherein absence yields lost production (i.e. absent workers cannot be replaced by existing unemployed individuals).[16] The costs of non-fatal road accidents are generated from two sources: temporary absence and permanent absence. Temporary absence applies to individuals that, due to a road accident, face a lower work capacity for a limited period of time, and therefore receive benefits from the SI. Permanent absence applies to individuals that, due to a road accident, have a permanent lower work capacity and thus receives DI benefits.

It should be noted that the HCA for estimating production losses has many limitations. The approach places no value on a retiree's temporary or lasting loss of the ability to work and does not value temporary disability among children, as they have not yet entered the labor force. Discounting future production losses to present value yields very low values for children and adolescents. Although they lose many years of work, those years are far in the future and heavy discounted. Furthermore, my estimates exclude production lost by people other than those injured as a result of an injury. These losses may include the time family, friends, and professionals spend caring for the injured, time spent investigating the injury, and worker retraining.

Costs are estimated using an incidence approach. This approach is based on the principle that the costs associated with an illness should be assigned to the year in which the costs begin. The production losses are measured as the present value of future income foregone by the individual due to a road accident and are assigned to the year in which the accident first occured. The total indirect

[16] The assumption that an absent worker's productivity is irreplaceable has been criticized as leading to an overestimation of the actual indirect cost. The idea is that individuals on short term absence can make up for lost production when they return, or their work tasks can be taken over by internal labor reserves, or that non urgent jobs can be cancelled. Individuals on long term absence can be replaced by someone currently unemployed. The cost of an illness or accident would thus be the friction costs; i.e. the search and training of new employees (Koopmanschap et al., 1995). For one of the purposes of this paper, investigating the cost of road accidents on the public sickness insurance system, this is not a problem. The SI and DI replace foregone earnings no matter if the claimant's production is replaced or not.

cost is thus determined by the individual employment rate, labor earnings, and, in case of permanent exit from the labor force, time until retirement. In the calculations, only the gross costs are included. Potential positive side effects that could be generated from, for instance, unemployed individuals replacing the absent workers are not taken into account (Hodgson & Meiners, 1982).

The present value formula I use to calculate the indirect costs is similar to what was suggested by Rice et al. (1989). Versions of this formula have also been used in previous Swedish cost of illness studies (see e.g. Henriksson & Jönsson, 1998; Jarl et al., 2008; MSB, 2009). Costs that accrue past the accident year are discounted 4 percent annually, according to recommendations by the Swedish Transport Analysis Agency (SIKA, 2009). A one percent annual growth rate is assumed. In order to facilitate the cost analysis is use the statistically significant estimates displayed in Table 3 and Table 4 are used to calculate the effects on days of sickness absence and disability insurance prevalence. The results from this excercise are presented in Table 5.

Table 5: Estimates of the change in days of sickness absence and disability insurance prevalence compared to the pre-accident period

	2004	2005	2006	2007	2008	2009	Total
Sickness Absence, days							
Men	12.79	10.35	4.87	2.39	-	-	30.40
Women	15.55	14.23	8.16	3.73	-	-	41.67
Total	13.71	11.64	5.96	2.84	-	-	34.15
Disability Insurance, prevalence,percent							
Men	-	1.05	2.24	3.40	4.22	4.27	
Women	-	-	3.58	6.23	6.81	6.47	
Total	-	0.71	2.68	4.35	5.08	5.00	

Note: Calculations are done based on the estimates controlling for individual characteristics. The effects are calculated as changes compared to the pre-accident period, 2002-2003.

I estimate the indirect costs of increased sickness absence as:

$$\sum_{t=1}^{4} PV_{SI}^{t} = D_{n,s,t} \frac{LE_{s,n}}{240} \frac{(1+g)^t}{(1+r)^t} \tag{5}$$

where PV_{SI} is the present discounted value of indirect costs from road accidents during the follow up period. $D_{n,s,t}$ is the number of sickness absence days due to road accident at year t for a person of age n and sex s. Age is divided in classes of ten years, and the median value of each class is used. $LE_{s,n}$ are the mean, annual labor earnings, including payroll taxes, by sex and age. I follow NCO (2008) and transform the annual labor earnings into daily earnings by dividing the production value by 242, the number of work days per year.[17] Finally, g is the increase in growth assumed to be 1 percent and r is the real discount rate assumed to be 4 percent.

The costs of increased DI up-take is calculated as:

$$\sum_{n=y}^{65} PV_{DI} = P_{y,s,n} LE_{s,n} \frac{(1+g)^{n-y}}{(1+r)^{n-y}} \tag{6}$$

Where PV_{SI} is the present discounted value of production loss from road accidents leading to permanent absence. $P_{y,s,n}$ is the probability that a person of a certain age, n, and sex, s, will survive to a certain age, y. As in equation 5, $LE_{s,n}$ are the mean annual labor earnings, including payroll taxes, for individuals of a certain sex and age. As before, g, the increase in growth, is assumed to be 1 percent and r, the discount rate is 4 percent.

As shown in section 4.1, road accidents victims are negatively selected. They have, on average, a lower employment rate and lower labor earnings compared to the popluation at large. If this is not taken into account, the cost will be overestimated. The average labor earnings by sex and age for the treatment group in 2003 - one year prior to the accident - will be used in the cost calculations. Included in labor earnings are all gross, cash compensation paid by employers for all individuals regardless of attachment to the labor market. Payroll taxes of 40 percent are added to the labor earnings.[18] Age is divided in classes of ten years, and the median value of each class is used.

Table B:1 in Appendix B presents the average labor earnings, with and without payroll taxes included, among road accident victims. Earnings are divided by gender and across age groups.

[17] 22 days per month for 11 months.

[18] Swedish payroll tax paid by the employer is usually around 30 percent of the employee salary. The percentage is lower for young and old employees. In addition, employers often pay 5 to 15 percent in fees to social insurances, according to collective agreements between employers and the union.

In order to calculate survival probabilities, I use Statistic Sweden's life tables. The probabilities are calculated as the ratio between the number of survivors at age n and the number of survivors at age y. Table B:2 in Appendix B shows the relevant data. I assume that the survival probabilites are unchanged following a road accident. This is clearly a strong assumption given the decline in health required to be granted DI.

7.1 Temporary absence

I start by calculating the indirect cost induced by temporary absence. As shown in section 6, a road accident is estimated to increase sickness absence by, on average, 34.15 (calendar) days over a six year period. The effect is stronger for women, 41.67 days, than men, 30.40 days. This amounts to a total of 239,143 days of sickness absence compensated by the SI. Men used 142,302 of these days and women used 96,841 days.[19] The majority of sickness absence occur during the year of the accident and the year after, but road accidents lead to additional sickness absence even three and four years after the event. Sickness absence is assumed to be evenly distributed across age groups. Table B:3 in Appendix B shows the total number of additional sickness absence days by year, gender, and age group.

Using equation 5, I calculate the value of the lost production due to temporary absence during each year of the follow up period. The result is presented in Table 6.

[19] The total amount of sickness absence, by gender and age group, is presented in Table B:3 in Appendix B.

Table 6: Estimates of the indirect cost due to temporary absence, by gender and year.

	2004	2005	2006	2007	Total
Men	12,076,765	9,217,146	4,211,852	2,007,382	27,513,147
Women	5,658,959	4,884,129	2,719,946	1,207,444	14,470,479
Total	17,735,724	14,101,275	6,931,799	3,214,826	41,983,626

Note: The indirect cost is estimated in terms of present value SEK in 2004. Costs that accrue past the accident year are discounted 4% annually. A one percent annual growth rate is assumed. The indirect costs are estimated using mean average earnings of the road traffic accident victims in 2003. Earnings are divided by gender and across age groups.

The total production value lost because of temporary sickness absence due to road accidents is estimated to be 42 million SEK (€4,6 million). Men are responsible for more than half of the production loss, 27.5 million SEK. This is partly due to the larger share of men among the victims, partly due to the fact that men, on average, have higher labor income than women, and partly due to the fact that they are affected more on average.

7.2 Permanent Absence

Table 5 shows that involvment in a road accident significantly increases the probability of receiving DI benefits. The entry into DI is growing over time. In 2006, two years after the accident, the additional DI take-up is 2.24 percent for men and 3.58 percent for women. The effect is estimated to increase to 4.27 percent, and 6.47 percent, respectively in 2009. In terms of individuals granted DI, this amounts to a total of 350 new DI claimants: 200 men and 150 women. The total productivity loss of permanent absence is calculated as the increased DI benefits taken up in 2009, the last year of the follow-up period. A potential drawback with this method is that claimants granted DI after 2009 are not included in the calculations. Given that the increase in DI appears to plateau in 2009, this is likely not a significant problem but may still lead to downward biased estimates.

As explained in section 3, DI benefits can be granted either fully or partially. Since I only have information about whether the individual receives DI benefits during the year or not, DI distribution due to a road accident is assumed to be equal to that of all newly granted DI cases.[20] The distribution across age groups is assumed to correspond to the distribution of newly granted DI categorized by the SIA as being due to "injury and poisoning etc." (Försäkringskassan, 2006). This gives me 270 full time DI beneficiaries: 116 women and 154 men. The

[20] In 2005, 58 percent of all new claimants were granted full time DI. Among those granted partial DI, 9 percent received it at 3/4, 64 percent received 1/2, and 27 percent were granted 1/4 DI (Försäkringskassan, 2006).

total amount of full time DI claimants, by gender and age group, is presented in Table B:4 in Appendix B.

Table 7 presents the estimated present value of the lost production due to permanent work absence following a road accident.

Table 7: Estimates of the indirect cost due to permanent absence, by gender and age group.

	Men	Women	Total
20 - 24	32,712,630	20,941,318	53,653,948
25 - 34	117,501,897	77,230,459	194,732,356
35 - 44	181,305,478	133,645,017	314,950,495
45 - 54	128,509,064	77,022,729	205,531,793
55 - 59	60,336,410	32,242,303	92,578,712
Total	520,365,479	341,081,826	861,447,305

Note: The indirect cost is estimated in terms of present value SEK in 2004. Costs that accrue past the accident year are discounted 4% annually. A one percent annual growth rate is assumed. The indirect cost are estimated using mean average earnings of the road traffic accident victims in 2003. Earnings are divided by gender and across age groups.

The total cost of permanent absence is estimated at approximately 861 million SEK (€95.7 million). Even though the relative effect on DI prevalence is larger for women, the majority of the induced cost is derived from men. As explained in section 7.1, this is partly due to a larger share of men among the treated and partly due to the fact that men, on average, have higher labor earnings.

The cost is largest for those aged 35 - 44, 315 million SEK (€35 million), and smallest for those aged 20 - 24, 53.6 million SEK (€5.9 million). The difference across age groups is a reflection of the number of new DI beneficiaries in each group and of the time to retirement: younger victims have more work years to lose compared to older ones. The last point becomes apparent if I instead look at the cost per newly granted DI. From this perspective the biggest loss comes from the age 20 - 24 group while the smallest cost is induced by road accident victims aged 55 - 59.

7.3 Cost per accident

The total present value of lost production is estimated at 900 million SEK (€100 million). In order to make the estimated cost comparable to the value used by the STA, I also calculate the estimated cost per road traffic accident. Following the method applied by the STA, I estimate the cost per road traffic accident to

142,008 SEK (€15,777).[21]

When compared to the value currently used by the STA, and what was suggested in MSB (2009), the estimated cost of road accidents in this study is significantly lower. I see two main reasons for this: (1) I am not able to account for all production losses. Only sickness absences of more than 14 days are identified. The majority of absence spells are shorter than that, and if they were included in the cost calculations, estimated costs would increase. Furthermore, only individuals of working age are included. This means that I am unable to include the cost of permanent absence from the labor market among children and adolescents in the estimates. (2) Previous studies have assumed that road traffic accidents are random in the population. To the individual, accidents may seem random and unpredictable but this study shows that this is not the case. In fact, there is a very clear negative selection into road accidents. I show that road accident victims, on average, are less educated, have a lower employment rate, and lower labor earnings than those who were not involved in a road accident.

As a sensitvity analysis, I test the last assumption by calculating the value of lost production using the mean average earnings of the control group instead of the road accident victim. As before, earnings are divided by gender and across age groups. The results from this exercise are presented in Table 8.[22] For comparison, I also include the baseline results from Section 7.1 and Section 7.2.

Table 8: Estimates of the indirect cost of road traffic accidents by treatment status

	Controls	Victims	Diff.
Total cost	1,079,387,822	903,430,931	175,956,891
Cost per accident	169,666	142,008	27,658

Note: The indirect cost is estimated in terms of present value SEK in 2004. Costs that accrue past the accident year are discounted 4% annually. A one percent annual growth rate is assumed. The indirect cost are estimated using mean average earnings of the control group and road traffic accident victims, respectively. Earnings are divided by gender and across age groups. Cost per accident is calculated from the net value of lost production, including VAT cost of 23 percent. Net value of the indirect costs are assumed to be 91 percent of the gross value.

I find that, under the assumption that road traffic accidents are random in the population, the estimated present value of lost production is estimated at approximately 1.1 billion SEK (€122 million) while the net production loss per accident is 170,000 SEK (€18,900). This is approximately 19 percent higher

[21] The cost per road accident is calculated as the net production loss, including VAT (total production loss * VAT (21 percent)) divided by the number of road accident victims. See Section 1 for a description and motivation of the methodology used.

[22] Appendix C presents the average labor earnings among the control group and a more detailed presentation of the temporary and permanent cost using the mean average earnings of the controls.

than what I find when I use the mean average earnings of road traffic accident victim's to estimate the lost production. This show that if the negative selection into road traffic accidents is not taken into account, the cost will be overestimated.

8 The Cost of Road Accidents to the Public Insurance System

As a way to reduce the number of road accidents, the idea of transferring the present costs of lost income from the obligatory and public sickness insurance to the traffic insurance has recently been discussed in a Swedish Government Official Report (SOU 2009:96). This would involve transfering money from the pool of individuals with private cars to the insured pool of employees. As described in Section 3 Swedish traffic insurance is obligatory but organized privately (around 40 companies are active on the market).[23] However, only a small part of the loss of income incurred by individuals injured in a road accident (accident victim) is compensated by the insurance. The major part of an accident victim's loss of income is instead compensated for by sick pay or sickness benefits, disability benefits, occupational injury insurance and occupational insurance schemes.

Transferring the cost to the traffic insurance would most likely increase the efficiency of the insurance and also reduce the number of accidents. The primary reason for increased efficiency is that private companies could affect the insured individual's behavior. For instance, one policy could be increasing the experience rating in the insurance (i.e. provide lower premiums for safe driving, driving safe cars etc.). Furthermore, in order to reduce the benefits paid from the insurance due to lost income, the insurer could provide more and better rehabilitation to accident victims. These incentives are not there today. Based on previous cost estimations, premiums paid to the Swedish traffic insurance would increase by approximately 70 percent as a consequence of making traffic insurance liable for all lost incomes (SOU 2009:96).

In this section I calculate the present value of the cost of road accidents to the Swedish public insurance system. As described in section 3, SI and DI replace forgone earnings due to temporary and/or permanent lower work ability. The SI replaces, on average, 80 percent of lost earnings, while the average DI replacement rate is 65 percent. Since SI and DI replace labor earnings, payroll taxes does not have to be taken into account in the calculations. The estimated costs of road accidents to the public insurance system are presented in Table 9. The costs are calculated using equation 5 and equation 6, but now $LE_{s,n}$ rep-

[23] The insurance can be seen as a "third party insurance" since it provides compensation in the event of being part (also as pedestrian or bicyclist) of a road accident.

resents 80 (65) percent of annual labor earnings, excluding payroll taxes.

The present value of the cost of road accidents to the public insurance system is estimated to be approximately 408 million SEK (€45 million). This is significantly less than the estimated production loss. In fact, the SI and DI only replace around 52 percent of lost production.

Table 9: The cost of road accidents to the public insurance system.

	SI	DI	Total
Men	15,721,832	232,924,132	52,948,468
Women	8,268,842	150,758,261	191,316,342
Total	23,990,674	383,682,393	407,673,068

Note: Cost is estimated in terms of present value SEK in 2004. Costs that accrue past the accident year are discounted 4% annually. A one percent annual growth rate is assumed.

9 Concluding Remarks

Road traffic accidents result in death, injury and property damage. For the individuals involved, the outcomes of a road crash may be devastating and it may not seem to them to be necessary or appropriate to place monetary values on these outcomes. However, determining the magnitude of this cost provides a better understanding of the benefits of activities that reduce the incidence and severity of road crashes. The cost of accidents are usually defined as direct costs (for health care, property damage, and administration), indirect costs (for lost production and consumption), and a value for safety per se (immaterial costs, i.e., pain, grief, suffering, joy of life etc.).

This paper uses detailed Swedish longitudinal administrative data to estimate the effect non-fatal road traffic accidents have on lost production. I utilize the fact that all Swedish workers are covered by public sickness and disability insurance and measure the production loss as changes in labor market absence following a road traffic accident. The costs are estimated as the present value of future earnings using an incidence approach. This means that all costs associated with the road accident are assigned to the accident year and future costs are discounted.

This study finds that the risk of being involved in a road accident is far from random. Individuals with a history of unemployment, sickness absence, social insurance uptake, and lower incomes are more likely to be involved in these accidents. These findings are in accordance with previous research suggesting that involvement in road traffic injuries is associated with lower socioeconomic status (see e.g. Graham et al. 2005; Laflamme & Engström 2002; Braver 2003; Hasselberg et al. 2005)

Road traffic accidents have dramatic and long-term effects on SI and DI utilization. The number of days on sickness absence increased, on average, by 34 days over a six year period. Men increased their sickness absence by 30.4 days while the effect for women was a 41.6 days increase. Five years after the accident, the increase in disability take-up rate was 5 percent. On average, the take-up rate increased by 4.27 percent for men and 6.47 percent for women. These long-term effects are in line with previous research on the impact of health shocks on future labor market outcomes (Dano, 2005; Lundborg et al., 2011; Halla & Zweimüller, 2013).

In economic terms, the cost of road accidents, measured as lost production, is estimated to be approximately 900 million SEK (€100 million), or 142,000 SEK per accident. For the public insurance system, the cost of road accidents is estimated to be around 410 million SEK (€45 million).

The estimated cost of road accidents in this study is significantly lower compared to what has been previously found (SIKA 2009, MSB, 2009). The main reason for this is: (1) I am not able to account for all production losses. Only sickness absences of more than 14 days are identified. The majority of absence spells are shorter than that. Furthermore, only individuals of working age are included. This means that I am unable to include the cost of permanent disability among children and adolescents in the estimates. (2) Previous studies have assumed that road traffic accidents are random in the population. This study shows that this is not the case. I show that if the negative selection into road traffic accidents is not taken into account, the indirect cost will be overestimated with around 19 percent.

References

Alfaro, J. L., Chapuis, M. & Fabre, F. (1994). *Socio-economic cost of road accidents: Final report of action COST 313*. Brussels, Commission of the European Community.

Altonji, J. G., Elder, T. E. & Taber, C. R. (2005). 'Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools'. *Journal of Political Economy*, 113(1), 151 - 184.

Andersson, H. (2005). 'The value of safety as revealed in the Swedish car market: An application of the hedonic pricing approach'. *Journal of Risk and Uncertainty*, 30(3), 211-239.

Angelov, N., Johansson, P. & Lindahl, E. (2013). *Gender differences in sickness absence and the gender division of family responsibilities* (No. 2013: 9). Working Paper, IFAU-Institute for Evaluation of Labour Market and Education Policy.

Bastida, J. L., Aguilar, P. S., & González, B. D. (2004). 'The economic costs of traffic accidents in Spain'. *Journal of Trauma and Acute Care Surgery*'. 56(4), 883-889.

Braver, E. (2003). 'Race, hispanic origin, and socioeconomic status in relation to motor vehicle occupant death rates and risk factors among adults'. *Accident Analysis & Prevention*, 35(3), 295 - 309.

Broström, G., Palme, M. & Johansson, P. (2002). *Economic incentives and gender differences in work absence behavior* (No. 2002: 14). Working Paper, IFAU-Institute for Labour Market Policy Evaluation.

Cedervall, M. & Persson, U. (1988). *Vägtrafikolyckornas personskadekostnader: en samhällsekonomisk beräkning av 1985 års personskadekostnader totalt och fördelat på åldersgrupper*. Lund

Connelly, L. B., & Supangan, R. (2006). 'The economic costs of road traffic crashes: Australia, states and territories'. *Accident Analysis & Prevention*, 38(6), 1087-1093.

Christens, P. (2001). *Sociale karakteristika hos trafikofre*. Notat 1:2001, Danmarks Transport Forskning.

Dano, A. M. (2005). 'Road injuries and long - run effects on income and employment'. *Health Economics*, 14(9), 955-970.

De Brabander, B., & Vereeck, L. (2007). 'Valuing the prevention of road accidents in Belgium'. *Transport Reviews*. 27(6), 715-732.
Elvik, R. (2000). 'How much do road accidents cost the national economy?'. *Accident Analysis & Prevention*, 32(6), 849-851.

Elvik, R., & Mysen, A. B. (1999). 'Incomplete accident reporting: Meta-analysis of studies made in 13 countries'. *Transportation Research Record: Journal of the Transportation Research Board*, 1665(1), 133-140.

Finkelstein, E. A., Corso, P. S., & Miller, T. R. (2006). *The incidence and economic burden of injuries in the United States*. Oxford University Press.

Försäkringsförbundet (2005), *Utvidgat ansvar för trafikförsäkringen*, Oktober 2005

Försäkringskassan (2006), *Nybeviljade sjukersättningar/ aktivitetsersättningar - statistik 2006:3*, Rapport 2006:3.

Graham, D., Glaister, S., & Anderson, R. (2005). 'The effects of area deprivation on the incidence of child and adult pedestrian casualties in England'. *Accident Analysis & Prevention*, 37(1), 125-135.

Halla, M., & Zweimüller, M. (2013). 'The effect of health on earnings: Quasi-experimental evidence from commuting accidents'. *Labour Economics*, 24, 23-38.

Hasselberg, M., Vaez, M., & Laflamme, L. (2005). 'Socioeconomic aspects of the circumstances and consequences of car crashes among young adults'. *Social Science & Medicine*, 60(2), 287-295.

Hauer, E. & Hakkert, A. (1988). 'Extent and some implications of incomplete accident reporting'. *Transportation Research Record*, (1185).

Henriksson, F. & Jönsson, B. (1998). 'Diabetes: The cost of illness in Sweden'. *Journal of Internal Medicine*, 244(6), 461 - 468.

Hodgson, T. A. & Meiners, M. R. (1982). 'Cost-of-illness methodology: A guide to current practices and procedures'. *The Milbank Memorial Fund Quarterly. Health and Society*, 429 - 462.

Hultkrantz, L., Lindberg, G., & Andersson, C. (2006). 'The value of improved road safety'. *Journal of Risk and Uncertainty*, 32(2), 151-170.

Jarl, J., Johansson, P., Eriksson, A., Eriksson, M., Gerdtham, U.-G., Hemström, R., Selin, K., Lenke, L., Ramstedt, M. & Room, R. (2008). 'The societal cost of alcohol consumption: an estimation of the economic and human cost including health effects in Sweden, 2002'. *The European Journal of Health Economics*, 9(4), 351 - 360.

Koopmanschap, M. A., Rutten, F. F., van Ineveld, B. M. & Van Roijen, L. (1995). 'The friction cost method for measuring indirect costs of disease'. *Journal of Health Economics*, 14(2), 171 - 189.

Laflamme, L., & Engström, K. (2002). 'Socioeconomic differences in Swedish children and adolescents injured in road traffic incidents: Cross sectional study'. *Bmj*, 324(7334), 396-397.

Larsson, J. (2008). *Trafikskador 1998-2005 enligt patientstatistik.* VTI rapport 615. Statens väg- och transportforskningsinstitut (VTI).

Lundborg, P., Nilsson, M. & Vikström, J. (2011). *Socioeconomic heterogeneity in the effect of health shocks on earnings: evidence from population-wide data on Swedish workers* (No. 2011:11). Working Paper, IFAU - Institute for Evaluation of Labour Market and Education Policy.

Mastekaasa, A. & Olsen, K. M. (1998). 'Gender, absenteeism, and job characteristics: A fixed effects approach'. *Work and Occupations*, 25(2), 195 - 228.

Mattsson, K. & Ungerbäck, A. (2013). *Vägtrafikolyckor: Handledning vid rapportering*. [Ny utg.]. Borlänge: Transportstyrelsen (STA).

Maraste, P., Persson, U., Berntman, M., & Svensson, M. (2002). *Kommunalekonomiska konsekvenser till följd av trafikolyckor: En åttaårsuppföljning av långvariga trafikskador*. Bulletin 208, Lunds Tekniska Högskola.

Maraste, P, Persson, U. & Berntman, M. (2003) 'Long-term follow-up and consequences for severe road traffic injuries—treatment costs and health impairment in Sweden in the 1960s and the 1990s', *Health Policy*, 66(2), 147-158.

Miller, T. R. (1997). Societal costs of transportation crashes. In *The Full Costs and Benefits of Transportation* (pp. 281-314). Springer Berlin Heidelberg.

Myndigheten för samhällsskydd och beredskap (2009). *Samhällets kostnader för vägtrafikolyckor: Beräkningar*. Karlstad: Myndigheten för samhällsskydd och beredskap (MSB).

Nationellt centrum för lärande av olyckor (2008). *Bränders samhällsekonomiska kostnader: Beräkningar*. Karlstad: Nationellt centrum för lärande av olyckor (NCO).

Nilsson, K., Persson, U. & Hjalte, K. (1997). *Kostnader för vägtrafikolyckor i Sverige och värdering av riskreduktioner: En översikt*. Borlänge: Vägverket.

Paringer, L. (1983). 'Women and absenteeism: Health or economics?'. *American Economic Review*, 73(2), 123 - 27.

Persson U., & Cedervall M. (1991). *The value of risk reductions: Results of a Swedish sample survey*. In: IHE Working Paper 1991, vol. 6.

Persson, U., Norinder, A., Hjalte, K., & Gralén, K. (2001). The value of a statistical life in transport: Findings from a new contingent valuation study in Sweden. *Journal of Risk and Uncertainty*, 23(2), 121-134.

Persson, U. & Vegelius, C. (1995). *Revidering av vägverkets olyckskostnader: en uppräkning till 1997 års priser*. Lund: Tekniska högskolan i Lund, Institutionen för trafikteknik.

Rice, D. P., MacKenzie, E. J., Jones, A. S., Kaufman, S. R., DeLissovoy, G. V., Max, W., ... & Smith, G. S. (1989). Cost of injury in the United States: A report to Congress.

Rosman, D.L. (2001). 'The Western Australian Road Injury Database (1987–1996): Ten years of linked police, hospital and death records of road crashes and injuries'. *Accident Analysis & Prevention*, 33(1), 81 - 88.

Statens institut för kommunikationsanalys (2009). *Värden och metoder för transportsektorns samhällsekonomiska analyser - ASEK 4*. Östersund: Statens institut för kommunikationsanalys (SIKA).

Segel, J. E. (2006). *Cost-of-illness studies—A primer*. RTI-UNC Center of Excellence in Health Promotion Economics, 1-39.

SFS 1975:1410. *Trafikskadelagen*. Stockholm: Justitiedepartementet.

Sjögren-Lindquist, G. & Wadensjö, E. (2006). *The Swedish welfare state: The role of supplementary compensations*. Working Paper Series 1/2006, Swedish Institute for Social Research.

Svensson, M. (2009). 'The value of a statistical life in Sweden: Estimates from two studies using the "Certainty Approach" calibration'. *Accident Analysis & Prevention*, 41(3), 430-437.

SOU 2009:96. *En utvidgad trafikförsäkring*. Stockholm: Socialdepartementet.

Trawén, A., Maraste, P., & Persson, U. (2002). 'International comparison of costs of a fatal casualty of road accidents in 1990 and 1999'. *Accident Analysis & Prevention*, 34(3), 323-332.

White, H. (1980). 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity'. *Econometrica: Journal of the Econometric Society*, 817 - 838.

Wijnen, W., Wesemann, P., & de Blaeij, A. (2009). 'Valuation of road safety effects in cost–benefit analysis'. *Evaluation and Program Planning*, 32(4), 326-331.

WHO (2004). *World report on road traffic injury prevention*. Technical report, World Health Organization, Geneva, Switzerland.

Appendix A Differences in Traffic Exposure

Table A:1: Summary of differences between road traffic accident victims and the control group, and group wise differences in traffic exposure according to The Swedish National Travel Survey, RES 2005-2006

Treated vs Controls	Exposure to traffic
Gender	
Treated: 67 % men	Men: 50 km / day
Controls: 51 % men	Women: 34 / day
Mean age	
Treated: 37.54	Older individuals travels further per day
Controls: 40.69	than younger individuals
Level of education	
Treated: 21 % upper secondary education	Small differences by level of education
Controls: 32 % upper secondary education	
Average labor income	
Treated: 152,520	Travel distance per day
Controls: 183,550	increases with income
Employment	
Treated: 69.3%	Employed: 46 km/day
Controls: 76.9%	Other (incl. unemployed): 23 km/day
Children in household	
Treated: 34%	Children: 44 km / day
Controls: 40%	No children: 40 m / day
Married	
Treated: 28.6%	Married: 45 km / day
Controls: 40.6&	Unmarried: 40 km / day

Source: Exposure to traffic according to The Swedish National Travel Survey, RES 2005-2006. The average kilometers travelled in traffic on a regular day. Modes of transport: walking, bike, MC, moped, bus, taxi, truck, and car. Ages 20 - 60.

Appendix B Values Used in Cost Estimations

Table B:1: Average annual labor earnings and production value for road accident victims. By gender and across age groups.

Age, y	Men		Women	
	Labor earnings	Production value	Labor earnings	Production value
20 - 24	102,326	143,257	69,794	97,711
25 - 34	166,840	233,575	113,645	159,103
35 - 44	182,587	255,621	132,319	185,247
45 - 54	186,798	261,517	169,027	236,638
55 - 59	180,044	252,061	157,815	220,941
60 - 64	159,910	223,873	126,261	176,766

Note: The table reports the average labor earnings and the production value for road accident victims the year prior to the accident. Labor earnings include all gross, cash compensation paid by employers. Production value is measured as labor earnings plus payroll taxes of 40 percent.

Table B:2: Survivors per 100 000 births, 2007 - 2011

	Men	Women
20 - 24	99,280	99,483
25 - 34	98,725	99,276
35 - 44	98.000	98,880
45 - 54	96,436	97,865
55 - 59	93,916	96,189
60 - 64	90,014	93,616

Source: Statistic Sweden, Population statistics, Life table. Survival rates are based on the median value of each age group.

Table B:3: The increase in temporary absence by age, gender and year

	Days of sickness absence					
Age	2004	2005	2006	2007	2008	2009
Men						
20 - 24	2.60	2.10	0.99	0.49		
25 - 34	3.44	2.78	1.31	0.64		
35 - 44	2.97	2.40	1.13	0.55		
45 - 54	2.43	1.97	0.92	0.45		
55 - 60	1.36	1.10	0.52	0.25		
Total	12.79	10.35	4.87	2.39		
Women						
20 - 24	2.70	2.47	1.42	0.65		
25 - 34	3.63	3.32	1.90	0.87		
35 - 44	3.63	3.32	1.90	0.87		
45 - 54	3.57	3.26	1.87	0.86		
55 - 60	2.03	1.86	1.07	0.48		
Totalt	15.55	14.23	8.16	3.73		
Full sample						
20 - 24	2.85	2.57	1.24	0.57	0	0
25 - 34	3.79	3.42	1.65	0.77	0	0
35 - 44	3.43	3.09	1.50	0.69	0	0
45 - 54	3.00	2.70	1.31	0.61	0	0
55 - 60	1.69	1.52	0.74	0.34	0	0
Total	14.77	13.30	6.44	2.99	0	0

Note:

Table B:4: Number of full time DI claimants due to a road accident, by gender and age

	Men	Women	Total
20 - 24	6	5	11
25 - 34	23	19	42
35 - 44	44	38	82
45 - 54	48	33	81
55 - 60	33	21	54
total	154	116	270

Notes: The number of new DI cases are based on all newly granted DI cases in 2005 (Försäkringskassan, 2006).

Appendix C Sensitivity Analysis: Using Earnings of the Non-victims.

Table C:1: Average annual labor earnings and production value for non-road accident victims. By gender and across age groups.

Age, y	Men		Women	
	Labor earnings	Production value	Labor earnings	Production value
20 - 24	87,071	121,899	67,939	95,115
25 - 34	189,329	265,060	125,983	176,376
35 - 44	244,136	341,790	162,918	228,085
45 - 54	247,867	347,014	183,252	256,552
55 - 59	230,151	322,211	165,602	231,843
60 - 64	180,297	252,416	126,661	177,325

Note: The table reports the average labor earnings and the production value for the control group the year prior to the accident. Labor earnings include all gross, cash compensation paid by employers. Production value is measured as labor earnings plus payroll taxes of 40 percent.

Table C:2: Indirect cost due to temporary absence, using mean average earnings of non-road accident victims. By gender and year.

	2004	2005	2006	2007	Total
Men	14,527,229	11,087,373	5,066,468	2,414,695	33,095,764
Women	6,319,817	5,454,502	3,037,584	1,348,451	16,160,354
Total	20,847,046	16,541,874	8,104,052	3,763,146	49,256,118

Note: Indirect cost is estimated in terms of present value SEK in 2004. Costs that accrue past the accident year are discounted 4% annually. A one percent annual growth rate is assumed. Mean average earnings of non-road accident victims are used in the calculations.

Table C:3: Production loss due to permanent absence, using mean average earnings of non-road accident victims. By gender and age group.

	Men	Women	Total
20 - 24	40,113,518	23,385,820	63,499,338
25 - 34	150,167,772	87,040,196	237,207,968
35 - 44	234,827,671	146,518,416	381,346,088
45 - 54	161,352,085	80,943,775	242,295,860
55 - 59	72,649,051	33,133,400	105,782,450
Total	659,110,097	371,021,607	1,030,131,704

Note: Indirect cost is estimated in terms of present value SEK in 2004. Costs that accrue past the accident year are discounted 4% annually. A one percent annual growth rate is assumed. Mean average earnings of non-road accident victims are used in the calculations.