

Are doctor's certificates worth while? Evidence from a social experiment*

Laura Hartman[♥], Patrik Hesselius[♣], and Per Johansson[♣]

Abstract

The paper exploits a unique Swedish social experiment carried out in 1988 to estimate the cost-efficiency of using doctor's certificates as a monitoring tool to verify worker absence. The treatment consists of postponing the requirement for a doctor's certificate from day eight to day fifteen. The experiment was conducted on 270,000 individuals in two geographical areas with the treatment group randomized by date of birth. The results for both regions show that extension of the waiting period has strong effects on sickness-absence duration. The experiment increased sickness benefits expenses but reduced the number of visits to a doctor. The net effect was clearly negative; costs exceeded benefits by a factor of five. Our estimated effect of monitoring on sickness absence is large and corresponds to a fully five percent increase in benefit size.

JEL codes: J22, J28, H55, I18

Keywords: Monitoring, Randomized Experiment, Absenteeism, Sickness Insurance

* Anders Forslund and Ulf Gabrielli are both gratefully acknowledged for bringing the experiment to our attention. We are also grateful to David Card, Eva Mörk, Per Pettersson-Lidbom and Olof Åslund for valuable comments as well as seminar participants at Uppsala University, Stockholm University, University of California, Berkeley and the Uppsala Workshop on Labour Market Effects of Social Insurance. Laura Hartman acknowledges financial support from the Wallander and Hedelius Foundation

[♥] (Earlier Larsson) Centre for Business and Policy Studies. e-mail: laura.hartman@sns.se

[♣] Institute for Labour Market Policy Evaluation (IFAU) and Department of Economics, Uppsala University. e-mail: patrik.hesselius@ifau.uu.se

[♣] Institute for Labour Market Policy Evaluation (IFAU) and Department of Economics, Uppsala University. e-mail: per.johansson@ifau.uu.se

1 Introduction

Coinsurance, co-payments, and deductibles together monitoring of eligibility rules are the main ways of fighting moral hazard within insurance. This applies to private insurance as well as public social insurances. The literature on the effects and importance of coinsurance and deductibles on social insurances is large; Kruger and Meyer (2002) provide a survey of these studies concluding that the level of coinsurance or benefit generosity plays a significant role in the labor-supply decisions of the insured. Studies from the US have mainly focused on worker's compensation and, most importantly, unemployment insurance, whereas the European literature mainly deals with unemployment insurance and public sickness insurance, which are common amongst European countries. Lately, the increasing use of disability insurance benefits on both sides of the Atlantic has attracted some attention among researchers; see e.g. Autor & Duggan (2006).

Empirical validation of the effects of monitoring lags behind the literature of benefit size. Most of the existing studies concern unemployment insurance. Meyer (1995) and Ashenfelter et al (2005) discuss results from US unemployment insurance experiments. The experiments studied in Meyer (1995) were designed to determine the effect on unemployment of cash bonuses or job-search programs that consist of both intensified job-search assistance and more extensive job-search verification. The experiments studied in Ashenfelter et al (2005) were designed to study the effect from stricter enforcement and verification of work search behavior. Ashenfelter et al (2005) conclude that job search verification does not provide any benefits. Based on the results from Meyer (1995) they conclude that job-search assistance may be beneficial for both the unemployed individuals and for the society. European evidence on the unemployment insurance is not as clear-cut, suggesting that intensified monitoring and sanctions indeed in some cases shortens unemployment duration.¹ Autor and Duggan (2006) and Karlstrom et al (2007) discuss the importance of eligibility rules within disability insurance systems.

¹ See for example Dolton & O'Neill (1996); van den Berg & van der Klaauw (2001); Blundell et al (2004); and Cockx & Dejemeppe (2007).

Our study estimates the effects of monitoring in the Swedish public sickness insurance. In more specific, we explore the cost-efficiency of doctor's certificate as a monitoring tool to verify the workers' eligibility to sickness benefits. Our contribution to the literature is twofold. First, and most importantly, there is no previous empirical evidence on the effect of monitoring within the sickness insurance. It is far from clear that the results from studies on other public insurance programs can be generalized to sickness insurance. The risk of sickness differs from the risk of unemployment, and thus also the individual's capacity to influence these risks. But sickness per se is not sufficient to qualify for sickness benefits. The condition is that the individual is too sick to work, while there is a formal search requirement in the unemployment insurance. Health (e.g. a mental disorder or a musculoskeletal disorder) is most likely more difficult to measure than formal search.² Hence, monitoring health conditions that prevent people from working is in many cases more difficult than monitoring the formal search effort. On the other hand, having a diagnosis of mental disorders is likely to be stigmatizing which should reduce the monitoring problem.

Second, randomized social experiments are rare in Europe which makes our data unique. The clean experiment design ensures that our results provide credible estimates of the causal effect on works absence from the requirement of medical certificates.

Section 2 outlines the experiment and the Swedish sickness insurance. The experiment was conducted on 270,000 individuals in two geographical areas (the Gothenburg municipal and Jämtland County) in the fall of 1988. Half of the eligible population (those born on an even date) was required to show a doctor's certificate at day fifteen in a work absence spell. For the other half a doctor's certificate was required from day eight in a work absence spell. In fact we have two different experiments, as the starting point varied among the regions. The initial requirement was from day eight in Gothenburg and from day fifteen in Jämtland. Thus, in one case treatment group experienced an increase of monitoring; in the other case the experiment implied a decrease of monitoring. This strengthens the external validity of our study.

² That is, information of health is to a large extent private and hence not surprisingly, there is a substantial literature on the validity of various health measures (see e.g., Aarts and de Jong (1992), Baker et al. (2004), Benitez-Silva et al., (1999, 2003, 2004), Bound (1991), Bound et al. (2001), Kerkhofs et al. (1999), Kreider (1999), Kreider and Pepper (2002, 2003), Kerkhofs and Lindeboom (1995) and Lindeboom and Kerkhofs (2003).

Section 3 presents the results. The fact that randomization was based on birth date allows us to reconstruct the treatment and control samples using administrative data (containing information on date of birth) from the National Social Insurance Board. The results are pronounced: the relaxation of monitoring significantly increases absence due to sickness. Among the treated, the length of sick spells increased by 0.64 days and by 0.59 days on average in the two areas respectively. The mean duration in the control group was approximately 15 days during the experiment period.

Section 4 assesses the economic importance of the results. The experiment increased sickness benefits expenses but reduced the number of visits to a doctor. The net effect, however, is that costs exceeded benefits by a factor of five. Furthermore, our estimated effect of monitoring on sickness absence corresponds to a fully five percent increase in benefit size.

These two measures lead us to conclude in Section 5 that the doctor's certificates are worth while to monitor eligibility for sickness insurance.

2 Swedish sickness insurance

Sweden has compulsory national sickness insurance. The system is financed by a proportional payroll tax and replaces earnings forgone due to temporary health problems that prevent the insured worker from doing her regular job. Almost all employed workers are automatically covered by SI.³ Benefits are related to the lost income during the sick spell.

Sickness benefits are, and have been, generous. In 1988, when the experiment was conducted, the vast majority of workers received 90 percent of their lost income from public insurance. A benefit cap excluded workers at the very top of the income distribution from receiving the full 90 percent. However, in addition to public insurance most Swedish workers are also covered by negotiated SI programs regulated in agreements between labor unions and employers' confederations. These insurance programs replace about 10 percent of forgone earnings, although there is considerable

³ SI also covers some non-employed individuals such as students and job-seekers registered with the Public Employment Service. Students usually receive the minimum amount of benefits whereas the sick benefits of the unemployed are determined by their income prior to unemployment.

variation between them. In this way, total compensation can in some cases reach 100 percent.

There is no limit on how often or for how long public insurance benefits are paid. Many sick spells continue for more than a year but examples of much longer durations exist. These long spells are more likely to lead to disability pensions than a return to work.

Because compensation levels are so high, one would expect monitoring of benefit claimants to be strict in order to reduce moral hazard. However, this has not been the case in Sweden. Specifically, a sick spell starts when the worker calls the public social insurance office to report sick. Within a week, at the very latest on the eighth day of sickness, the claimant must verify eligibility by showing a doctor's certificate that proves reduced working capacity due to sickness. The public insurance office then judges the certificate and decides upon further sick-leave. It is very rare that the certificate is *not* approved. During 2006 the request for sick pay was rejected in 1.5 per cent of all new sick spells (SSIA, 2007).

Of course, some exemption rules make it possible for the public insurance offices to monitor more (or less) strictly. When abuse is suspected, they may visit the claimant at home.⁴ Claimants who have been on sickness benefits too frequently in the past may be asked to show a doctor's certificate from day one.⁵ Moreover, a new sick spell starting within five working days of the first is counted as a continuation of the first, making it impossible to report sick every Monday without ever visiting a doctor. Individuals with chronic illnesses, on the other hand, need not verify their eligibility each time illness forces them to remain at home.

Since 1988, some features of the system have changed. The compensation level has been reduced to 80 percent. The benefit cap has not been fully inflation-adjusted and as a result, in 2003, one quarter of the workforce currently receives less than 80

⁴ We have no statistics on the prevalence of home visits, however based on discussion with administrators at the SSIA we believe that this ever happened in practice in 1988. Monitoring of suspected abuse is at date performed only by matching registers and target against individuals with criminal offenses.

⁵ Unfortunately there is no statistic available on the prevalence of when SSIA is/was requiring a certificate from the first day. From the discussion with administrators and caseworkers and at the SSIA we believe this is/was very rare. The caseworkers we talked with had no own experience. In the Swedish journal for doctors ("läkartidningen") there is a discussion on a change in the rules from January 2008 where also employers (as from 1991 there is an employer responsibility in the insurance) were allowed to require doctors certificates for the first day one can see that most doctors were not aware of the possibility that the SSIA also could require a certificate for the first day (see <http://www.lakartidningen.se/engine.php?articleId=6993>).

percent of their forgone earnings from public insurance. In 1993, deductible was further increased by an uncompensated qualifying day at the beginning of each sick spell. Since 1991, some of the financial (and administrative) burden has been shifted to employers by requiring that they pay sickness benefits during the first weeks of sickness.⁶

The monitoring of sickness absence is thus very light. It consist basically in the requirement of adoctors certificate as of the eight day. It has been though bee shown that doctors find it difficult to carry out their function as a gatekeeper and that they to a large extent only agree with the patient requirement on sicklistning (see e.g., Arrelöv, 2006 and Englund, 2008). The meeting at day sevn can thus more be cosdeed as *transaction cost for the worker and not as a verification of illness that prevent people from working.*

3 The experiment

The experiment we use to identify the effect of monitoring was carried out in the second half of 1988 in *Gothenburg*, the second largest city in Sweden, and in *Jämtland*, a small county in the sparsely populated Northern part of Sweden. It was initiated by the local social insurance offices.⁷

The purpose of the experiment was to assess whether and how sickness absence behavior changed when monitoring of insurance claimants was reduced. A randomly assigned treatment group was allowed to receive sickness benefits for two weeks without showing a doctor's certificate, instead of one week as usual. The randomization was based on date of birth. All insured born on an even date were assigned to the group with two weeks unmonitored sickness absence whereas insured born on an uneven date had to adhere to the normal rules (one week unmonitored sickness absence).

The insurance authorities provided several arguments for running the experiment, all based on the presumption that it would result in savings and less sickness absence. First, eliminating unnecessary doctors visits would reduce costs for individuals, the medical care system and thereby the state budget. Second, the implementing authorities also believed that doctors systematically prescribed longer than necessary absences from work. With a two week gratis period, more individuals would return to work before receiving such a prescription. Finally, some sick spells were indeed expected to

⁶ Employers' responsibility for sickness benefits, or 'sick pay', has varied from 2 to 4 weeks. As of June 2005, employers pay full benefits for two weeks of sickness (except for the qualifying day) after which they bear a small share of the total cost for the remainder of the sick spell.

⁷ Until recently, the public insurance system was administered by 21 independent local social insurance offices that were free to allow exceptions from the general rules, as long as the exception leant toward more generosity. Today, administration is centralized.

get longer as truly sick individuals were not prematurely forced back to work. This in turn would decrease the risk of recurrence amongst those individuals.

The background and starting point of the experiment differed between the two areas. The idea to test the two-week rule for issuance of a doctor's certificate originated in Jämtland, where it had been in use for *all* insured since January 1987. Before 1987, the normal one week unmonitored sickness period was used in Jämtland. In Gothenburg, the usual rule of one week's unmonitored sickness absence was in use until the experiment. Thus, in Gothenburg, the experiment implied looser rules for half of the insured, whereas in Jämtland it implied harder rules for half of the insured. In spite of this, to make the presentation clear we label the group with a two-week waiting period as the treatment group and the group with a one week waiting period as the control group.

The experiment was a non-blind experiment in that all participants were informed in advance or during the experiment. The experiment was preceded by comprehensive local information campaigns and, in addition to personnel at local social insurance offices, employers and medical centers were informed in advance about the experiment. Written material such as brochures and posters were distributed, as well as verbal information through meetings and consultants. The mass media were also an important channel to inform the insured. Furthermore, a brief summary of the experiment was printed on the form that every insured reporting sick must fill in to receive sickness benefits.

To reconstruct the treatment and control samples and to evaluate the effect of the experiment, we use data from the National Social Insurance Board. The data include detailed information about individual sick spells, date of birth and numerous other characteristics. *Table 1* shows the distribution of insured individuals in the treatment and control groups in the two experiment areas. The control group is as expected, larger than the treatment group as there are more uneven than even birth dates and the rates also correspond to the expected 179/365.25 and 186.25/365.25.

Table 2 presents descriptive statistics for Gothenburg and Jämtland subdivided into the control group and treatment group respectively. There are no significant differences between the treatment and control groups with respect to their characteristics. Gender and age distributions, as well as average age and average

Formatiert:
Kursiv

sickness absence prior to the experiment, are indistinguishable between the treatment and the control group. Thus, the randomization seems to be valid.

A striking observation from *Table 2* is that sick spells were more common in Gothenburg than in Jämtland prior to the experiment. Recall that in Jämtland all insurance claimants were covered by a two-week rule since January 1987.

To further explore the validity of the randomization device we take a closer look at sickness absence among the treatment and the control groups prior to the experiment.

Figure 1 and 2 show the estimated survival functions in the pre-experimental period, i.e. the fraction of individuals still absent from work after a certain number of days into a sickness spell.⁸ As expected, there is no significant difference between the treatment and the control groups.

Both figures do, however, indicate a monitoring effect. In Gothenburg, *before the experiment*, we observe a drop in the survival rate around the eighth day of a sick spell. The pre-experimental survival functions in Jämtland look predictably different: the drop in the survival function is displayed on the 15th day instead of the eighth day.

4 The results

We begin by looking at how monitoring affects absence duration. Given the shapes of the survival functions in Gothenburg and Jämtland prior to the experiment we have a strong prior: the looser the checks on eligibility, the longer the absence spells. Besides duration, we are also interested in the incidence of sick spells. We address heterogeneity by estimating the effects for various types of individuals.

4.1 Duration

Figure 3 and *4* use survival functions to illustrate how monitoring affects the length of the sick spells. The dotted line indicates the fraction of ongoing sick spells among the treated and the unbroken line shows the corresponding fraction among the controls. The effect is distinct: the survival rate is significantly higher for the treatment group throughout the second week.⁹

⁸ We use the Kaplan-Meier estimator throughout the survival analysis (Kaplan and Meier, 1958).

⁹ The survival estimates and corresponding 95 percent confidence intervals are presented in Appendix A.

Hazard rates are a related way to analyze the effect of monitoring and offer a more detailed picture of when the probability of ending a sick spell is largest. *Figure 5* and *6* show the results for Gothenburg and Jämtland.¹⁰ It is evident that the control group – monitored at day eight – tends to exit after seven days of absence, whereas the treatment group waits another week before returning back to work. The hazard rates also exhibit a weakly cyclical pattern with a peak in the hazard rate every seventh day. This is probably due to a common practice among medical doctors to put patients on a sick-list a week at a time.

Furthermore, the figures indicate that the effect of monitoring is observable some days before the ultimate verification date. The hazard rate of the control group lies above that of the treatment group as early as days five and six. The opposite holds for days twelve and thirteen: the probability of returning back to work is higher among the treated. We label this a “pre-monitoring effect” and the main part of this effect is due to weekends. Most employees work Monday to Friday and have the weekend off. Thus, in the register, a sickness spell that starts on a Monday often ends on a Friday, even though the total absent from work is 7 days.

There also appears to be a “post-monitoring effect”. The hazard rates of those just monitored are below the hazard rate of the comparison population for some time after the date of monitoring. This may potentially be due to sample selection, or a *harvesting effect* according to epidemiology literature (see e.g. Schwartz, 2000). Let us be more specific. It is reasonable to assume that the date of monitoring “pushes” some people back to work who would otherwise have remained longer on sick-leave. When forced to show a doctor’s certificate, they perceive the cost of staying on sick-leave higher than the cost of returning to work sick. If so, the average health of the remaining individuals still on sick leave is worse than the average health of the comparison population immediately after the date of monitoring. Consequently, their hazard rate is lower on those days than it would have been without monitoring. An alternative explanation is that doctors prescribe longer sick-leaves than necessary, as discussed in Section 2. In this case, the post-monitoring effect could be viewed as a ‘causal’ determinant of the monitoring device.

¹⁰ Hazard rates are estimated using the product-limit method.

As a way to summarize the treatment, pre- and post-treatment effects in one measure, we estimate the mean duration given a follow-up time of three to three and a half years.¹¹ The mean estimates are presented in column one of *Table 3*. The estimates are presented separately for Jämtland and Gothenburg and for the treated and control group, respectively.¹² Based on these means we find a fairly large and statistical significant effect in Gothenburg, whereas a small and statistical insignificant effect is found for Jämtland. These results clearly differs from the results in *Figures 3* and *4* where an almost identical monitoring effects in the two areas was seen. The reason for the difference in results is the influence of long duration (or outliers) on the mean estimate. A few single spells of more than three years in sickness absence have a large effect on the uncensored means, especially if sample size is small.¹³ Since the population is smaller in Jämtland than in Gotheburg is the risk of influential observations (or outliers) higher in Jämtland than in Gothenburg. To accommodate this problem we decompose the mean into two components: (1) the mean duration of all spells followed for a maximum of four weeks and (2) expected remaining duration for work absence spells that lasted more than four weeks.

In column two of *Table 3* we can see that the full effect of the reform is for SI durations shorter than four weeks (column 2). The implication of this analysis is that relaxing the non-monitoring period from seven to fourteen days yields a significant increase in duration by 0.64 days in Gothenburg and by 0.59 days in Jämtland. In relation to mean duration in the control group, the duration was increased by 4.2 percent in Gothenburg and by 3.7 percent in Jämtland.

regression

4.2 Incidence

One of the arguments for running the experiment was to reduce the risk of recurrence by preventing the insured from being forced back to work 'too early' while still ill. If such an effect dominated, it would imply fewer sick spells during and after the experiment

¹¹ The mean duration is obtained by integration of the survival curve from infinity to zero.

¹² Note that percentiles for the treated and controls can be obtained directly from *Figures 3* and *4*. From these figures we can see that all action is for the 50 – 20 percentile. The difference between the tread and controls for the 20:th percentile is almost seven days.

¹³ There was at the time of the experiment no formal (nor informal) time limits on the length of a sickness spell.

period. *Figure 7* and *8* display the incidence of sickness spells in Gothenburg and in Jämtland before, during and after the experiment. Estimates of each half-year's incidences including the standard errors are presented in *Table 4*. In short, we find no statistically significant difference in sickness incidence between the control and experimental individuals before, during and after the social experiment.

4.3 Are the effects the same for everybody?

The above results demonstrate that the effect of monitoring is not statistically different in Gothenburg as in Jämtland. This is a reassuring result for several reasons. First, it is convincing evidence that the control group did not receive the treatment.¹⁴ Recall that the starting points were different in these two areas. In Jämtland, the two-week rule had been applied to *all* insured since January 1987 whereas in Gothenburg, the usual rule of one week's unmonitored sickness absence was in use until the experiment. We would thus expect the risk of the control group being treated to be higher in Jämtland than in Gothenburg, simply because it is harder to implement more stringent than looser monitoring. This in turn would imply that the estimated effect in Jämtland would be more biased towards zero than the effect in Gothenburg.

It also gives us grounds to dismiss concerns about the so-called Hawthorne effect, in which individuals respond to an experiment because they know that their behavior is being measured. Specifically, asymmetry in starting points could imply different Hawthorne effects in Gothenburg and in Jämtland. Comparing the survival functions during the experiment (*Figure 3* and *4*) with the pre-experimental survival functions (*Figure 1* and *2*), we find that behavior is insensitive to starting points. The treated appear to behave like all insured in Jämtland did before the experiment, and the controls behave like all insured in Gothenburg did before the experiment.

Finally, and perhaps most importantly, the similarity of the effects across regions makes it possible to generalize our results. The labor market in Jämtland is very different to that in Gothenburg. Jämtland is a rural area in Northern Sweden whilst Gothenburg is Sweden's second largest city and located in the south west. The average level of education is higher in Gothenburg than in Jämtland and thus the white- to blue-

¹⁴ See Moffitt (2004) for a nice discussion of shortcomings with external validity in randomized trials.

collar worker ratio is much higher in Gothenburg than in Jämtland. Because the estimated effects in the two regions are so similar, despite their labor market differences, we believe that the average effect can be generalized to Sweden overall.

Nonetheless, we are still interested in other potential dimensions of heterogeneity. Anecdotal evidence suggests, for example, that morale is lower among the young than the elderly. Besides differences according to age, we also test whether the effect of monitoring differs with respect to gender and income.

In order to, primarily, simplify the presentation we estimate proportional hazard models (cf. Lancaster, 1990) for each category and report the relative risk of monitored group relative of the non-monitored group at day seven and fourteen, respectively. Estimating proportional hazard instead of empirical hazard is not that restrictive since there is random assignment to treatment and control. That is, the standard critique of erroneous non-proportionality between the two groups is not relevant. However we do restrict the treatment effects to occur on days seven and fourteen in a sickness spell, thus neglecting the pre- and post-treatment effects.

Table 5 summarizes the results. A relative risk ratio above (below) one means that the group monitored at that date has a larger (smaller) probability of exit than the group not monitored that date. The first row reports the hazard estimates for all observations, also shown in *Figures 5* and *6*. The following rows present hazard estimates by gender, age and income. The only statistically significant difference in the monitoring effect is between men and women - the relative risk of the monitored is higher among men than among women both at day seven and fourteen. Men appear to react more strongly to monitoring than women do. This result also holds when controlling for age and income. In contrast, neither age nor income level plays a significant role. The estimated relative risks are indistinguishable across age and income groups. In other words, the effect of monitoring appears to be insensitive to these characteristics.

How should we explain this difference in behavior between the sexes? One explanation is that men simply have more lax moral. However, we cannot rule out that selection may also play a role. The intuition for this is as follows. Individuals who are absent due to sickness likely exhibit different characteristics to the overall labor force. Women are absent due to sickness more often than men, which suggests that the ‘sick’ female population *on average* differs less from the labor force than does the ‘sick’ male

population. This implies, in turn, that ‘sick’ women behave differently than ‘sick’ men. Another potential explanation for women’s more frequent sick absence is that child-bearing and -rearing lead to more frequent and better contact with medical services, lowering the indirect cost of visiting a doctor relative to men.

It is worth noting that in all subgroups the effect of monitoring is somewhat larger at day seven than at day fourteen. We believe that this is due to the *harvesting effect* discussed earlier: at day fourteen, the sample of controls consists of people with worse health than the treated. Thus by day fourteen, the two groups are not longer unconfounded. However, *Figures 5 and 6* indicate that this bias is small and the results in *Table 5* suggest it is not statistically significant. The monitoring effects at days seven and fourteen are statistically significantly different at the 5 percent level only when estimated for all observations in Gothenburg. When estimated for any of the subgroups, or for all observations in Jämtland, the difference is no longer statistically significant.

5 Economic implications

We have shown that monitoring plays a statistically significant role in explaining the duration of compensation claims. The sample sizes are large why statistical significance may not be of such large policy value. To make inference about the *economic significance* of monitoring we introduce two measures. First, we compare the costs and benefits of the experiment. Second, we compare the estimated effect of monitoring to the effect of an alternative reform, namely an increase of the compensation level.

5.1 Costs and benefits of relaxing monitoring

Our simple cost-benefit analysis compares the public cost of longer sickness duration to the public benefit of fewer visits to a doctor.¹⁵ Under the (tested) assumption that the incidence into sickness absence is unaffected by the extended waiting period the increased cost from experiment is easily obtained. The cost of the experiment is then obtained by first estimating the increased number of days in work absence for the treated population. This resulting figure is then multiplied by the average cost of one

¹⁵ Health care in Sweden is heavily subsidized and mainly provided by the counties.

day absence. The number of sick spells for the second half year in 1987 for the treated population was 121,963. The increase in sickness absence duration was estimated to be 0.64 days and in 1988. Thus the experiment implied a total increase in work absence by 73,177 days. The average cost of one day absence (in terms of sickness benefits) was at the date of the experiment 296 Swedish kronor (SEK).¹⁶ Taking these figures together the cost of the experiment was approximately 23 million SEK.

The benefit of the reform is calculated as the average cost of a visit to a doctor *times* the decrease in the number of visits.¹⁷ According to estimates from the Federation of Swedish County Councils (Landstingsförbundet), the average cost of a doctor's visit in 1988 was 445 SEK of which the patient paid a fee of 50 SEK.¹⁸ Thus, the public cost of a doctor's visit was 395 SEK. The reduction in the number of visits is again obtained from our estimated survival functions (see Table A1 in Appendix A). In the control group, 21.2 percent of all spells were longer than one week and thus involved at least one visit to a doctor. In the treatment group, 11.9 percent of the spells were longer than two weeks. Thus extending the waiting period to fourteen days would at maximum reduce the number of visits to a doctor by 9.3 percentage points.¹⁹ This gives that an extension of the waiting period by a week for all insured would reduce the number of doctor's visits by 11,343. The inferred benefit of 4.5 million SEK is considerably lower than the estimated cost of 23 million SEK.

An alternative way to present the simple cost-benefit calculation is to divide the total cost of 23 million SEK by the decreased number of potentially forced physician visits due to the certification requirement (11,343). This yields the result that the reform would have been cost neutral if the average public cost of a physician visit was 2,034 SEK. Compared to the actual public cost of 395 SEK, one can easily see that the cost of the reform greatly exceeded the benefit.

¹⁶ In 1988, US\$ 1 corresponded to 6.1 SEK.

¹⁷ We thus neglect benefits from less administrative costs for the social insurance offices. These benefits are likely to be small however.

¹⁸ The cost of a visit to a general practitioner in 1988 is calculated using the cost in 1991 deflated by the average increase in the visit cost for internal medicine and ear, nose and throat care between 1988 and 1991 (Landstingsförbundet, 1988, and Landstingsförbundet, 1991)

¹⁹ It is likely that a reasonable fraction of those absent due to sickness saw a doctor for medical treatment before the certificate requirement needed to be enforced. Thus, the actual reduction in the number of doctor's visits is probably less than 9 percentage points.

5.2 Monitoring and the level of compensation

Insurance theory tells us that moral hazard can be reduced either by increasing monitoring or by reducing compensation. This raises an interesting policy question: by how much must the level of compensation be increased to yield the same increase in work absence caused by the one-week relaxation of monitoring?

Using the estimated hazard rates for the two groups, we simulate the proportional increase of the hazard rate necessary to obtain an expected duration equivalent to monitoring on day fifteen. The calculations presented in Appendix B show that monitoring on the fifteenth day instead of at day eight is equivalent to a 1.4 percent proportional decrease of the hazard rate. Johansson & Palme (2005) estimates an increase in the compensation level by one percent would decrease the hazard from work absence by 0.25 percent. Under the assumption of no duration dependence, we find that a fully five percent increase in the compensation level is needed to increase duration in sickness absence by the same extent as introducing one-week relaxation of monitoring. In terms of the compensation level in 1988, this implies increasing the compensation level from 0.90 to 0.95.

6 Conclusions

We have shown that the degree of monitoring plays an important role in reducing moral hazard in the Swedish sickness insurance system. This evidence is based on a well-performed randomized controlled experiment that is not only internally valid but also highly externally valid.

Results from this experiment, together with results from previous studies on excess in sickness insurance, suggest that postponing the point of monitoring by one week corresponds to an increase in the compensation level of fully 5 percent. From a policy perspective this is an important trade off: the distributional (and thus equity) effects of increased monitoring are considerably different from the distributional effects of an overall reduction in the compensation level.

An interesting aspect of the study is the heterogeneity between men and women in sickness-absence behavior. Monitoring seems to have a stronger effect on men than on

women. On the other hand, women are absent due to sickness more often than men. Thus, *men as a group* are hit harder by increased monitoring, whereas *women as a group* are hit harder by an increase in excess. Of course, in order to judge the implications for equity it is important to know why sickness absence is higher among women. To our knowledge, this remains a question for future research.

References

Arrelöv B (2006). Läkarna i sjukskrivningsprocessen i *SKA Projektet: Sjukförsäkring, kulturer och attityder*. Edward Palmer (red), *Analyserar* 2006:16, Stockholm: Försäkringskassan.

Ashenfelter, O., Ashmore, D., Deschêns, O., 2005. Do Unemployment Insurance Recipients Actively Seek work? Randomized trials in Four U.S. States. *Journal of Econometrics*, 53-75

Atkinson, A., Micklewright, J., 1991. Unemployment compensation and labour market transitions: A critical review. *Journal of economic literature* 29, 1679-1727.

Carling, K., Holmlund, B., Vejsiu, A., 2001. Do benefit cuts boost job findings? Swedish evidence from the 1990s. *Economic Journal* 111, 766-790.

Chiappori, P., Salanié, B., 2000. Testing contract theory: a survey of some recent work. Invited lecture, World Congress of the Econometric Society in Seattle, 2000.

Dolton, P., O'Neill, D., 1996. Unemployment duration and the restart effect: Some experimental evidence. *Economic Journal* 106, 387-400.

Englund L (2008). Hur har distriktsläkares sjukskrivningspraxis förändrats under 11 år? Resultat av tre praxisundersökningar bland distriktsläkare i ett svenskt landsting åren 1996, 2001 och 2007. Arbetsrapport: Centrum för Klinisk Forskning Dalarna, Falun.

Ercolani, M., Barmby, T., Treble, J., 2002. Sickness Absence: An International Comparison. *Economic Journal* 112, F315-F331.

Henreksson, M., Persson, M., 2004. The Effects on Sick Leave of Changes in the Sickness Insurance System. *Journal of Labor Economics* 22(1), 87-113.

Hesselius, P., 2004. Sickness Absence and Labor Market Outcome. *Economic Studies* 82, Department of Economics, Uppsala University.

Holmlund, B., 1998. Unemployment insurance in theory and practice. *Scandinavian Journal of Economics* 100, 113-41.

Johansson, P., Palme, M., 1996. Do Economic Incentives Affect Work Absence? Empirical Evidence Using Swedish Micro Data. *Journal of Public Economics* 59, 195-218.

- Johansson, P., Palme, M., 2002. Assessing the Effects of a Compulsory Sickness Insurance 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, 1879-1890.
- Kaplan, E., Meier, P., 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53, 457-481.
- Krueger, A., Meyer, b., 2002. Labor supply effects of social insurance. in Auerbach and M Feldstein (eds.) *Handbook of Public Economics* 4. Amsterdam: North-Holland.
- Lancaster, T., 1990. *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.
- Larsson, L., 2004. Harmonizing unemployment and sickness insurance: Why (not)? *Swedish Economic Policy Review* 11(1), 151-188.
- Larsson, L., 2006, Sick of being unemployed? Interactions between unemployment and sickness insurance. *Scandinavian Journal of Economics* 108(1), 97-113.
- Landstingsförbundet, 1988. Kostnader per intagen patient, vård dag, läkarbesök m.m. 1988 (Cost per patient, day of care, doctor's visits etc 1988).
- Landstingsförbundet, 1991. Kostnader per intagen patient, vård dag, läkarbesök m.m. 1991 (Cost per patient, day of care, doctor's visits etc 1991).
- Meyer, B., 1995. Lessons from the U.S. Unemployment Insurance Experiments. *Journal of Economic Literature* 33(1), 91-131
- Moffitt, R. A., 2004. The role of randomized field trials in social science research. *American Behavioral Scientist* 47, 506-540.
- Schwartz, J., 2000. Harvesting and long term exposure effects in the relation between air pollution and mortality. *American Journal of Epidemiology* 151, 440-448.
- SSIA (2007). Nej till sjukpenning vad hände sen? Försäkringskassan Analyserar 2007:1.
- van den Berg, G., van der Klaauw, B., 2001. Counseling and monitoring of unemployed workers: Theory and Evidence from a social experiment. Working paper 2001:12, IFAU, Uppsala.

Appendix A

Table A1. Survival rates

Time	Gothenburg						Jämtland					
	Control			Treatment			Control			Treatment		
	Fraction remaining	95% KI	Difference									
1	100	100 - 100	0.0	100	100 - 100	0.0	100	100 - 100	0.0	100	100 - 100	0.0
2	85.4	85.2 - 85.5	0.2	85.6	85.4 - 85.8	0.2	81.5	81.0 - 81.9	0.3	81.7	81.3 - 82.2	0.3
3	69.5	69.2 - 69.7	0.5	69.9	69.6 - 70.2	0.5	64.1	63.5 - 64.6	0.4	64.5	63.9 - 65.1	0.4
4	58.4	58.1 - 58.6	1.0	59.4	59.1 - 59.7	1.0	52.7	52.1 - 53.2	0.7	53.4	52.8 - 54.0	0.7
5	50.1	49.8 - 50.4	1.5	51.6	51.3 - 51.9	1.5	44.6	44.0 - 45.1	1.4	46.0	45.4 - 46.6	1.4
6	39.3	39.0 - 39.5	3.9	43.1	42.9 - 43.4	3.9	35.6	35.1 - 36.2	2.8	38.5	37.9 - 39.0	2.8
7	34.5	34.2 - 34.7	4.5	39.0	38.7 - 39.3	4.5	31.4	30.9 - 31.9	3.5	34.9	34.3 - 35.5	3.5
8	21.2	20.9 - 21.4	12.4	33.6	33.3 - 33.9	12.4	20.3	19.9 - 20.8	10.0	30.4	29.8 - 30.9	10.0
9	19.7	19.5 - 19.9	10.6	30.2	30.0 - 30.5	10.6	18.5	18.1 - 19.0	9.0	27.5	27.0 - 28.0	9.0
10	18.9	18.7 - 19.2	9.0	27.9	27.7 - 28.2	9.0	17.7	17.2 - 18.1	7.7	25.3	24.8 - 25.9	7.7
11	17.9	17.7 - 18.1	7.6	25.6	25.3 - 25.8	7.6	16.6	16.2 - 17.1	6.6	23.2	22.7 - 23.7	6.6
12	16.9	16.7 - 17.1	6.5	23.4	23.2 - 23.7	6.5	15.6	15.2 - 16.1	5.8	21.4	20.9 - 21.9	5.8
13	15.7	15.5 - 15.9	4.7	20.4	20.1 - 20.6	4.7	14.5	14.1 - 14.9	4.5	19.0	18.6 - 19.5	4.5
14	14.6	14.4 - 14.8	4.0	18.6	18.4 - 18.8	4.0	13.5	13.2 - 13.9	4.0	17.5	17.1 - 18.0	4.0
15	12.9	12.7 - 13.1	-1.0	11.9	11.7 - 12.1	-1.0	12.1	11.7 - 12.4	-0.4	11.7	11.3 - 12.1	-0.4
16	12.2	12.0 - 12.3	-0.8	11.4	11.2 - 11.6	-0.8	11.4	11.0 - 11.7	-0.1	11.3	10.9 - 11.6	-0.1
17	11.7	11.6 - 11.9	-0.7	11.1	10.9 - 11.2	-0.7	10.9	10.6 - 11.3	0.1	11.0	10.7 - 11.4	0.1
18	11.1	10.9 - 11.2	-0.4	10.6	10.5 - 10.8	-0.4	10.4	10.0 - 10.7	0.3	10.6	10.3 - 11.0	0.3
19	10.4	10.3 - 10.6	-0.2	10.2	10.0 - 10.4	-0.2	9.9	9.6 - 10.3	0.3	10.2	9.8 - 10.5	0.3
20	9.8	9.7 - 10.0	-0.1	9.7	9.5 - 9.9	-0.1	9.4	9.1 - 9.8	0.3	9.7	9.3 - 10.0	0.3
21	9.3	9.1 - 9.4	0.0	9.2	9.1 - 9.4	0.0	9.0	8.7 - 9.3	0.3	9.3	9.0 - 9.7	0.3
22	8.4	8.3 - 8.6	0.1	8.6	8.4 - 8.7	0.1	8.5	8.2 - 8.8	0.3	8.8	8.5 - 9.1	0.3
23	8.2	8.0 - 8.3	0.1	8.3	8.1 - 8.4	0.1	8.2	7.9 - 8.5	0.3	8.5	8.2 - 8.8	0.3
24	8.0	7.9 - 8.2	0.1	8.1	8.0 - 8.3	0.1	8.1	7.8 - 8.4	0.2	8.3	7.9 - 8.6	0.2
25	7.7	7.5 - 7.8	0.1	7.8	7.6 - 7.9	0.1	7.7	7.4 - 8.0	0.3	8.0	7.7 - 8.3	0.3
26	7.4	7.2 - 7.5	0.1	7.5	7.3 - 7.6	0.1	7.5	7.2 - 7.8	0.3	7.7	7.4 - 8.0	0.3
27	7.1	7.0 - 7.2	0.1	7.2	7.1 - 7.4	0.1	7.2	6.9 - 7.5	0.2	7.4	7.1 - 7.8	0.2
28	6.8	6.7 - 7.0	0.1	6.9	6.8 - 7.0	0.1	7.0	6.7 - 7.3	0.2	7.1	6.8 - 7.4	0.2

Appendix B

Equal sickness absence rate

Above we saw that the duration in work absence was expected to be longer for the population with a fourteen days waiting period instead of the population with a seven days waiting period. The expected duration was 0.6 days longer on average. In this appendix we derive how one can obtain the proportionate adjustment to the hazard that would yield the same effect on expected duration as was estimated by the experiment.

First, note that the survival function in discrete time can be expressed as

$$(B1) \quad S(t) = e^{-\sum_{i=0}^{t-1} h(i)},$$

where $h(t)$ is the hazard rate at time t . Furthermore, the expected duration is given by

$$(B2) \quad E[T] = \sum_{t=1}^{\infty} S(t).$$

From (B1) and (B2) it is evident that the expected duration with a proportionally-adjusted hazard rate $((1+p)h(i) \quad \forall i)$ can be expressed as

$$(B3) \quad E[T_p] = \sum_{t=1}^{\infty} S_p(t) = \sum_{t=1}^{\infty} e^{-\sum_{i=0}^{t-1} (1+p)h(i)} = \sum_{t=1}^{\infty} \left[e^{-\sum_{i=0}^{t-1} h(i)} \right]^{(1+p)} = \sum_{t=1}^{\infty} S(t)^{(1+p)}$$

The expected duration of the treatment (T) and control (C) group can be written

$$(B4), (B5) \quad E[T_C] = \sum_{t=1}^{28} S_C(t) + \sum_{t=29}^{\infty} S_C(t) \quad \text{and} \quad E[T_T] = \sum_{t=1}^{28} S_T(t) + \sum_{t=29}^{\infty} S_T(t).$$

Here the first terms are the contribution to the expected duration for the first 28 days of survival, while the second term is the remaining contribution.

For our sample the first term for the control group (monitoring on day eight) is estimated to 6.84 while for the treatment group (monitoring at day 15) we have an estimate of 7.44. For simplicity, assume these estimates to be the population parameters then we get

$$(B6) \quad E[T_C] = \sum_{t=1}^{28} S_C(t) + \sum_{t=29}^{\infty} S_C(t) = 6.84 + \sum_{t=29}^{\infty} S_C(t)$$

$$(B7) \quad E[T_T] = \sum_{t=1}^{28} S_T(t) + \sum_{t=29}^{\infty} S_T(t) = 7.44 + \sum_{t=29}^{\infty} S_T(t).$$

Since the experiment did only affect spells during maximum the first 28 days the last terms in (B6)(B6) and (B7)(B7) are set to be equal. Thus,

$$(B8) \quad \sum_{t=29}^{\infty} S_C(t) = \sum_{t=29}^{\infty} S_T(t) = \sum_{t=29}^{\infty} S(t) = 8.19.$$

By using (B3)(B3) the expected duration for the control group with a proportionally-adjusted hazard rate $E[T_p]$ can be expressed as

$$(B9) \quad E[T_p] = \sum_{t=1}^{\infty} S_C(t)^{(1+p)} = \sum_{t=1}^{28} S_C(t)^{(1+p)} + \sum_{t=29}^{\infty} S(t)^{(1+p)}$$

Setting the expected duration of the proportionally-adjusted control-group's hazard rate equal to that of the mean duration in the treatment group, 15.63, we get

$$(B10) \quad E[T_p] = \sum_{t=1}^{28} S_C(t)^{(1+p)} + \sum_{t=29}^{\infty} S(t)^{(1+p)} = 15.63.$$

It is possible to numerically solve for p . We obtain $p = -0.014$. Thus, a 1.4 percent proportional decrease of the hazard will yield the same aggregated sickness absence as the extension of the monitoring from seven to fourteen days.

Feldfunktion

Feldfunktion

Feldfunktion

Table 1. Number of insured individuals, 1 July to 31 December, 1988.

	Control group	Treatment group	Total
Jämtland county			
Non-government employees	30,221	28,978	59,199
Government employees	5,646	5,450	11,096
All employees	35,867	34,428	70,295
Gothenburg			
Non-government employees	106,825	102,603	209,428
Government employees	16,956	15,928	32,884
All employees	123,781	118,531	242,312

Table 2. Descriptive statistics.

	Gothenburg		Jämtland	
	Control group	Treatment group	Control group	Treatment group
No. individuals	106,825	102,603	30,221	28,978
Percent women	50.4 (0.15)	50.3 (0.16)	50.9 (0.29)	50.6 (0.29)
Mean age	38.0 (0.04)	37.9 (0.04)	38.8 (0.07)	38.9 (0.08)
Yearly wage (SEK) ¹ :				
Mean	118,579 (169.05)	119,260 (173.06)	100,820 (249.68)	101,023 (259.97)
Percent above benefit cap	6.8 (0.08)	7.0 (0.08)	2.0 (0.08)	2.1 (0.09)
Sickness absence January 1, 1988 to June 30, 1988:				
Incidence	1.10 (0.004)	1.10 (0.005)	0.92 (0.007)	0.90 (0.008)
Prevalence ²	16.61 (0.12)	16.61 (0.12)	14.20 (0.20)	13.90 (0.20)
Average duration ³	15.1	15.1	15.4	15.4

Note: Standard errors in parentheses. ¹ The estimated non-absence yearly income of which the sickness absence benefit is based on. SEK 100 corresponds to US \$12.80 in June 2005. ² Average number of absence days during the studied period. ³ Assumes equilibrium (inflow equals outflow) and is calculated by dividing the prevalence by the incidence.

Table 3. Expected remaining duration.

	Mean duration, followed max 3 to 3,5 years	Mean duration followed max. 28 days
Gothenburg		
Treated	16.14 (0.220)	7.57 (0.025)
Nr. censored		
Control	15.27 (0.212)	6.93 (0.024)
Nr. censored		
Difference	0.87 (0.306) [0.0045]	0.64 (0.034) [<0.001]
Jämtland		
Treated	15.82 (0.460)	7.13 (0.052)
Nr. censored		
Control	15.73 (0.463)	6.54 (0.049)
Nr. Censored		
Difference	0.10 (0.652) [0.88]	0.59 (0.071) [<0.001]

Notes: Standard errors in parentheses, p-values of testing the null hypothesis of difference equal zero in brackets. A spell starting in the second half of 1988 is followed maximum until December 31st 1991 due to data limitations. Thus, a spell starting in July 1988 has a follow-up period of 3.5 years and a spell starting in December 31st 1988 has a follow up period of 3 years.

Table 4. Incidence.

	Pre-treatment half-year 1/1/88 – 6/30/88	Treatment half-year 7/1/88 – 12/31/88	Post-treatment half-year 1/1/89 – 6/30/89
Gothenburg			
Treated	1.10 (0.004)	1.19 (0.005)	1.11 (0.005)
Control	1.10 (0.005)	1.19 (0.005)	1.11 (0.005)
Difference	-0.002 (0.0063) [0,71]	-0.005 (0.0064) [0.39]	-0.003 (0.0064) [0.63]
Jämtland			
Treated	0.90 (0.007)	0.98 (0.008)	0.95 (0.008)
Control	0.92 (0.008)	0.98 (0.008)	0.97 (0.008)
Difference	-0.021 (0.011) [0.044]	-0.008 (0.011) [0.46]	-0.018 (0.011) [0.12]

Note: Standard errors in parentheses, p-values of testing the null hypothesis of difference equal zero in brackets.

Table 5. Estimated relative risks between the monitored and the unmonitored groups at day 7 and 14.

	Gothenburg				Jämtland county			
	Relative risk ^a		Parameter		Relative risk ^a		Parameter	
	Day 7	Day 14	Day 7	Day 14	Day 7	Day 14	Day 7	Day 14
All observations	3.24	3.58	1.176 (0.014)	1.277 (0.024)	3.11	3.48	1.133 (0.033)	1.248 (0.054)
Gender:								
Men	3.87	4.38	1.353 (0.021)	1.478 (0.035)	3.88	4.32	1.356 (0.048)	1.464 (0.079)
Women	2.73	2.88	1.003 (0.020)	1.058 (0.033)	2.55	2.69	0.935 (0.045)	0.988 (0.075)
Age:								
16-25	3.66	4.04	1.296 (0.032)	1.396 (0.052)	3.82	3.91	1.341 (0.059)	1.362 (0.088)
26-35	3.05	3.38	1.115 (0.027)	1.218 (0.044)	2.91	3.66	1.070 (0.055)	1.298 (0.096)
36-45	3.17	3.51	1.155 (0.027)	1.257 (0.045)	2.74	2.76	1.008 (0.070)	1.016 (0.116)
46-55	3.28	3.37	1.188 (0.034)	1.216 (0.058)	2.58	3.23	0.949 (0.106)	1.172 (0.183)
56-65	2.99	4.33	1.094 (0.079)	1.465 (0.160)	2.81	0.78	1.034 (0.339)	-0.254 (0.606)
Income:								
1 st Quartile	3.46	3.66	1.242 (0.029)	1.298 (0.051)	2.54	3.80	0.931 (0.067)	1.335 (0.129)
2 nd Quartile	3.29	3.91	1.192 (0.026)	1.363 (0.044)	3.30	3.16	1.194 (0.064)	1.152 (0.105)
3 rd Quartile	3.04	3.41	1.112 (0.030)	1.227 (0.050)	3.14	3.41	1.146 (0.067)	1.228 (0.106)
4th Quartile, below cap	3.19	3.29	1.160 (0.038)	1.192 (0.061)	3.34	3.48	1.205 (0.085)	1.246 (0.123)
Over cap	3.23	3.20	1.173 (0.051)	1.163 (0.079)	3.66	4.01	1.296 (0.107)	1.389 (0.160)

Note: Standard errors in parentheses. ^a Relative risk = exp(parameter).

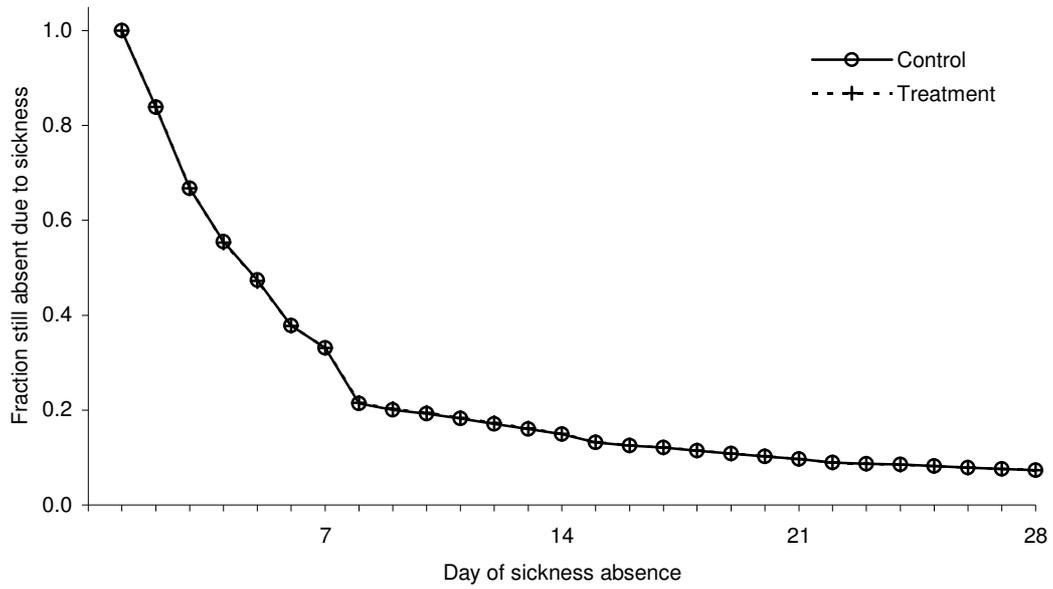


Figure 1. Fraction still absent due to sickness in Gothenburg during the half year before the experiment period (1/1/88 – 6/30/88).

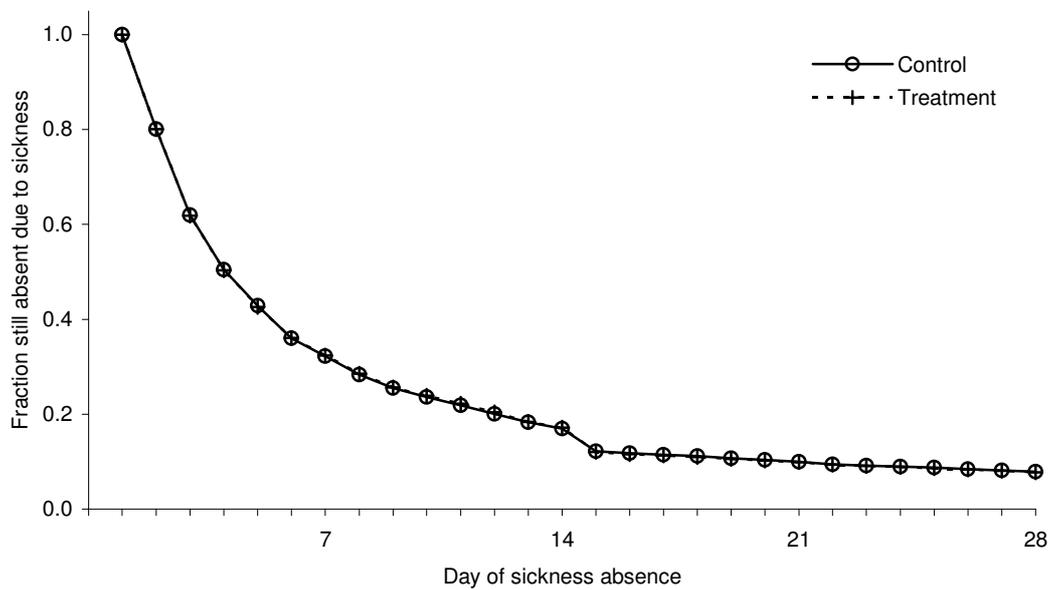


Figure 2. Fraction still absent due to sickness in Jämtland during the half year before the experiment period (1/1/88 – 6/30/88).

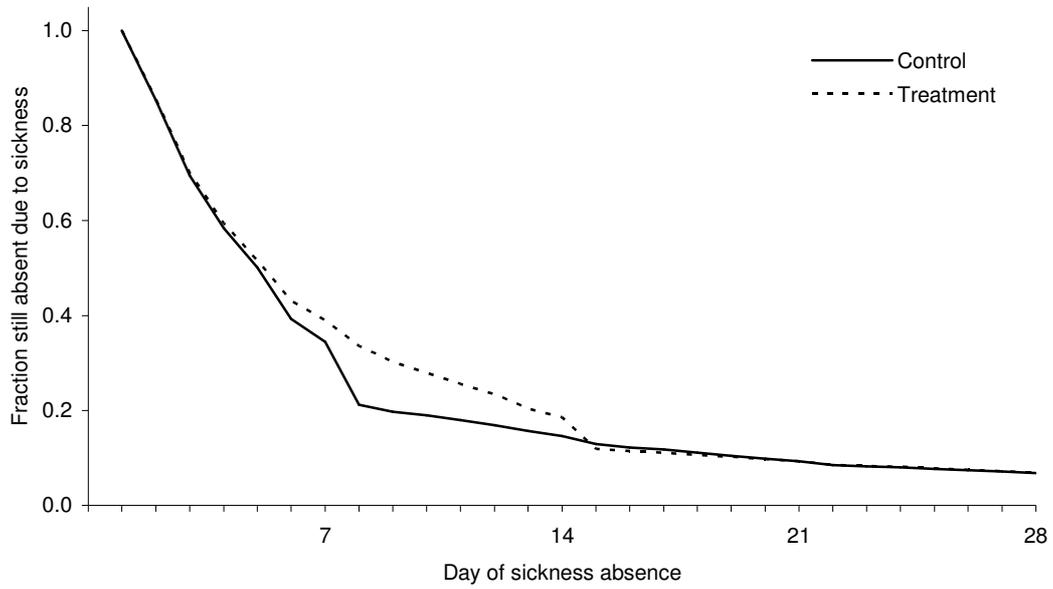


Figure 3. Fraction still absent due to sickness in Gothenburg during the experiment period.

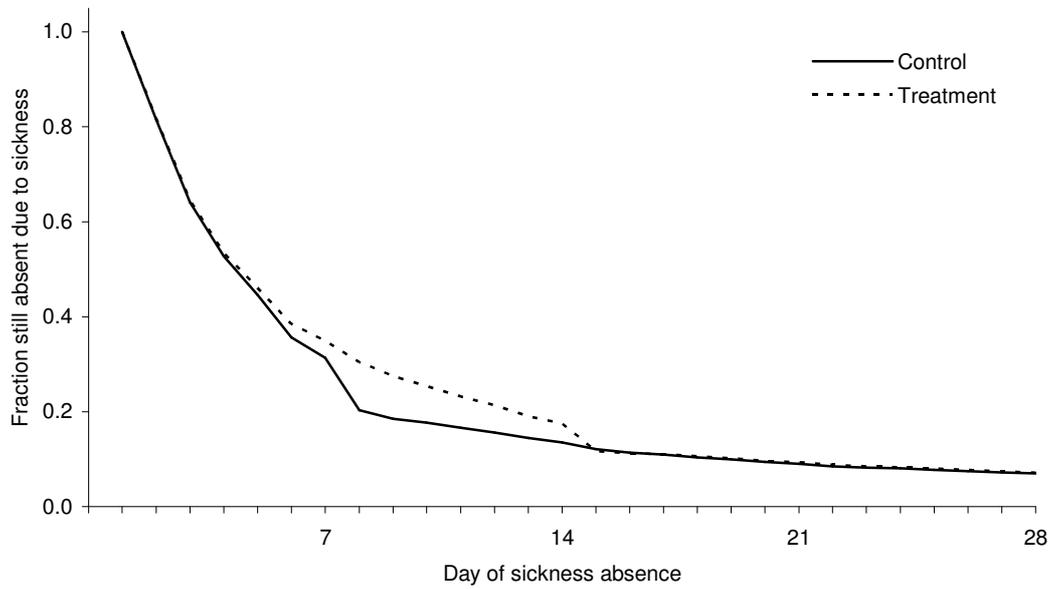


Figure 4. Fraction still absent due to sickness in the county of Jämtland during the experiment period.

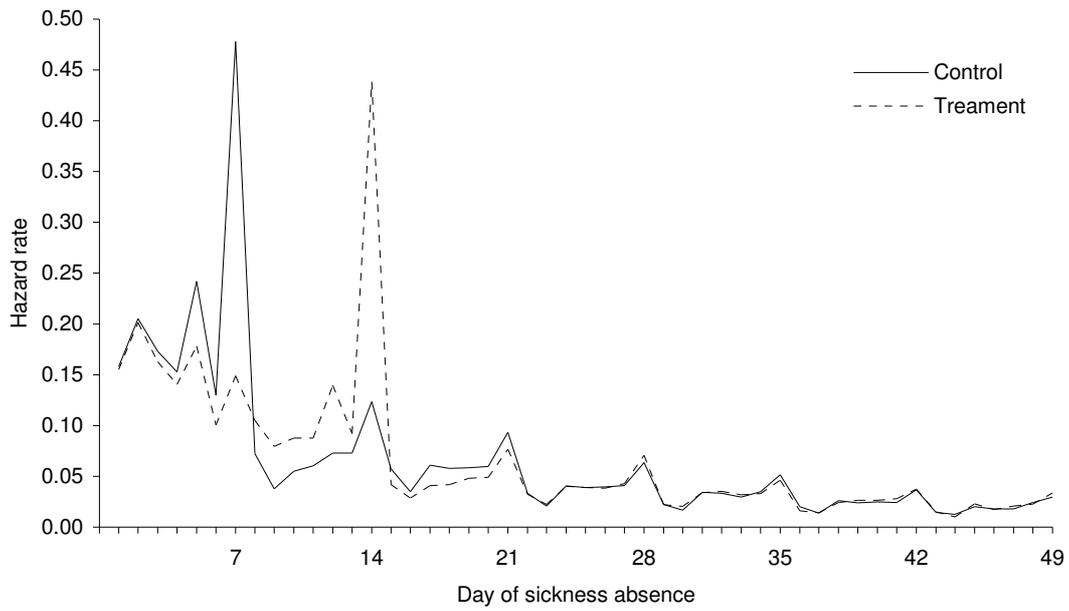


Figure 5. Hazard rate during the period of the experiment, Gothenburg.

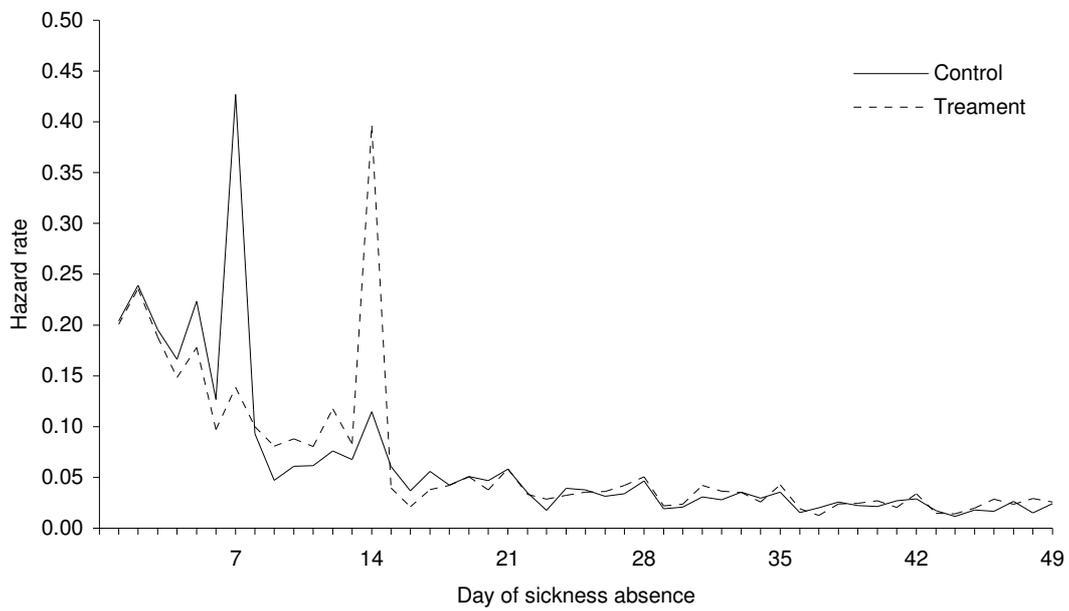


Figure 6. Hazard rate during the period of the experiment, county of Jämtland.

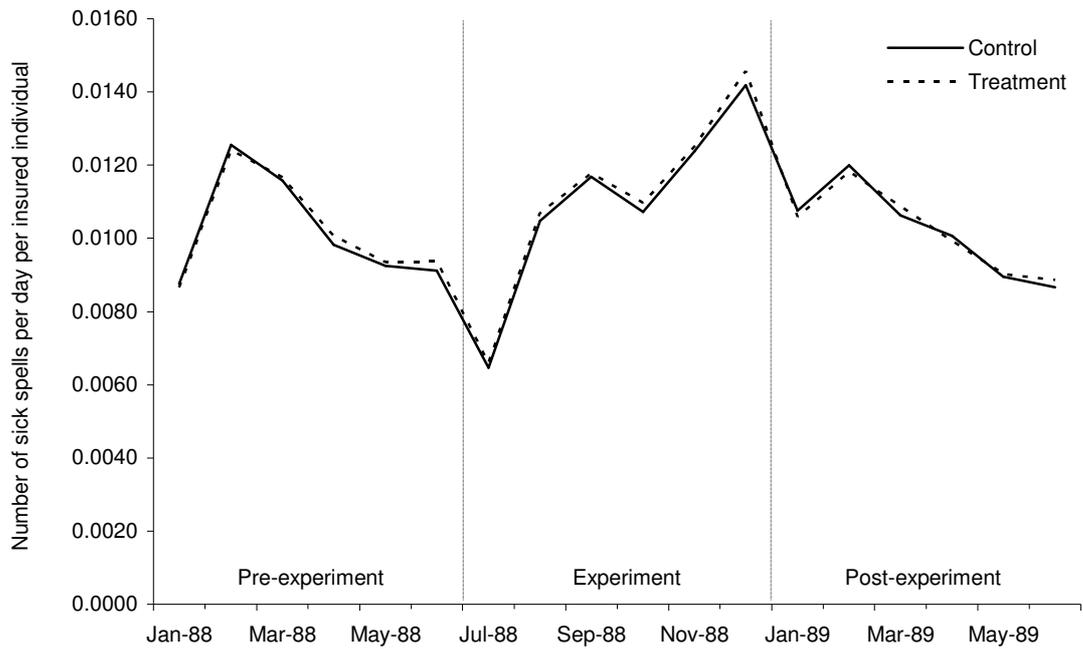


Figure 7. Sickness absence incidence among insured individual in Gothenburg, before, during and after the experiment.

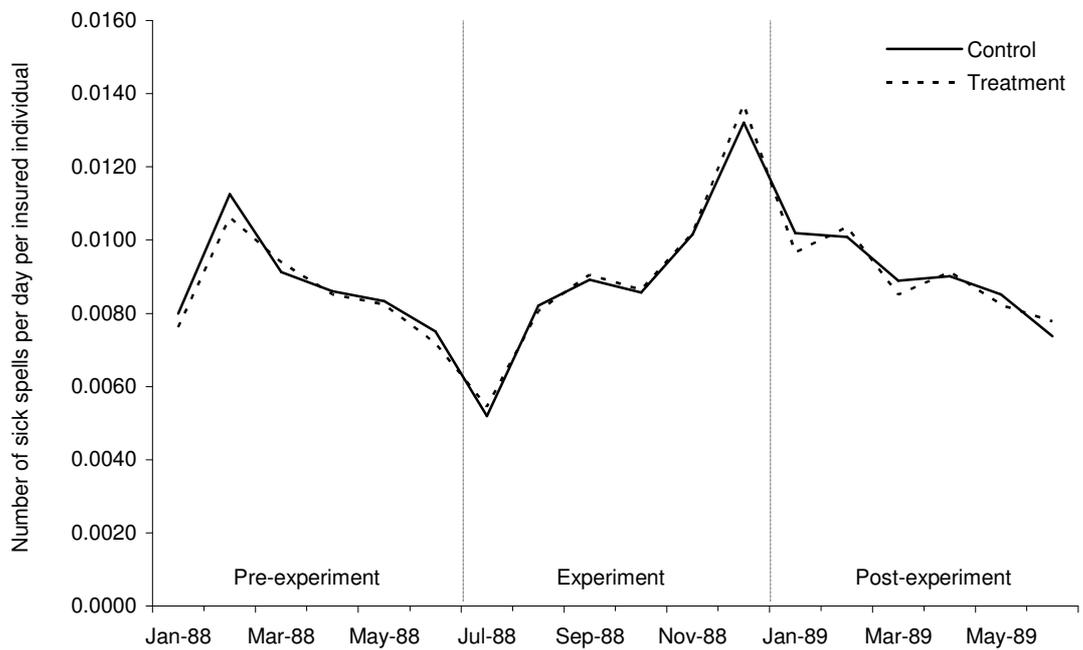


Figure 8. Sickness absence incidence among insured individual in Jämtland, before, during and after the experiment.