The individual cost of sick leave

Simen Markussen[1]

Ragnar Frisch Centre for Economic Research

January 2010

Abstract:

This paper assesses the causal effect of sick leave on subsequent earnings and employment, using an administrative dataset for Norway to link individual earnings, sick-leave records and primary care physicians. To create an experiment-like setting where similar workers are given different sick-leave durations, the leniency of a worker's physician - certifying sickness absences - is used as an instrumental variable for sick leave. I find that a one percentage point increase in a worker's sick-leave rate reduces his earnings two years later by 1.2 percent. Around half of the reduction in earnings can be explained by a reduction of 0.5 percentage points in the probability of being employed.

Keywords: sickness absence, wage formation, IV estimation, wage regression

JEL classification: C31, J22, J24, J31, J33, J71

[1] This paper is part of the project "A viable welfare state" financed by the Norwegian Research Council. I am grateful to Knut Røed, Oddbjørn Raaum, Bernt Bratsberg, Espen Henriksen, Ole Røgeberg, Erik Biørn, Kjell Arne Brekke, Magne Mogstad, Harald Dale-Olsen and Per Johansson. A previous version of this paper is available in the Memoranda series at the University of Oslo, Department of Economics (No. 20/2009).

1. Introduction

There's no such thing as a free lunch. Despite that Norwegian workers enjoy full wage replacement during sickness; sick leave is not for free. Two years later, workers with sick leave earn less and are more seldom employed than others. This paper estimates the individual cost of sick leave using Norwegian register data covering all workers, their primary care physicians and the sick-leave spells certified by these physicians. To isolate causal effects of sick leave from other factors, an instrumental variable strategy is applied. The practice style, or leniency, of each primary care physician is used as an instrumental variable for individual sick leave.

Generally, IV-regression captures the effect for a "specific group of people – namely, people whose behavior can be manipulated by the instrument. (…) That is, the effect is estimated for subjects who will take the treatment if assigned to the treatment group, but otherwise not take the treatment" (Angrist and Krueger, 2001, p.77). There is an important discretionary component related to sickness absence. Whether a worker with a certain health problem is able to work or needs sick leave partly depends on subjective judgment. It is this judgment that is captured by the instrumental variable used in this paper. The estimates should be seen as a comparison of two similar workers with the same health problem but with different sick-leave durations due to differences in their physicians' judgment.

I find that sick leave has a substantial impact on earnings two years later. On average, a one percentage point increase in an individual's sick-leave rate reduces this person's earnings two years later by 1.2 percent. Taking into account the fact that earnings are fairly persistent over time, these effects make it relevant to ask whether sickness insurance really is insurance – or if it is a loan against future earnings. A one percentage point increase in sick leave also reduces the probability of being employed full-time two years later by 0.5

percentage points. Consequently, the effects on earnings are partly caused by reduced employment.

Physicians' leniency, used as instrumental variable, is estimated in conjunction with a rich set of observables for each worker (earnings, education, age, family situation, county of residence etc.) including a workplace dummy, and is taken from Markussen et al. (2009). Physicians' leniency, if well measured, should be a suitable instrumental variable for sick leave, capturing the variation in absence propensity due to differences in leniency among doctors. A series of robustness tests supports this. First, a comparison of physicians' leniency with physicians' quality, measured as age and gender-adjusted mortality rates per physician, suggests that differences in *leniency* across physicians are unrelated to differences in *quality*. Second, by controlling for more than 13 000 census tracts (grunnkrets), variations related to geography that can potentially feed into the leniency indicator are eliminated. Third, the potential selection bias arising from *non-random matching* of workers and physicians are tested by estimating the costs of sick leave on a subsample consisting only of workers who have been assigned to a new physician, and where this assignment is outside of their control. Fourth, a potential problem of *reflection* arises as the leniency indicator in essence is a group mean in which the individual worker contributes. The importance of this is tested by randomly dividing the patients of each physician into two groups, estimating a new leniency indicator based on the first group's sick leave, and using this new indicator as an instrumental variable for sick leave in the second group. The estimated results turn out to be robust to all these tests.

The estimated effects differ among different workers. First of all: Men's earnings are much more affected by sick leave than what is the case for women. This can partly be explained by a gender-segregated labor market. In jobs dominated by women, their sick leave has almost no effect on their earnings. Interestingly, the effects on employment also seem smaller in these jobs. Women in jobs dominated by men are "treated like men" in the sense that the effects of sick leave on earnings are fairly similar for both genders. In these jobs, the effects of sick leave on employment for women are stronger than in jobs dominated by women. A possible explanation for this is that different employers offer their workers different implicit contracts. Some employers offer "low absence/high pay" contracts while others offer "high absence/low pay" contracts. This reflects that in some jobs low sick-leave rates (high reliability) are compensated with high wages, and this may be an important factor for understanding why the labor market is gender segregated.[2]

Generally, the findings indicate that wage effects are strongest at the top of the income distribution (or education distribution) while employment effects are strongest at the bottom. The results also indicate that the effect of sick leave on earnings is non-linear such that the marginal cost of sick leave on earnings decreases with spell duration. If we consider sick leave a measure of productivity or effort, the highest possible effort is when absences are zero. Hence, sick leave is a censored signal, and starting an absence spell removes the worker from the high effort / zero absence group.

Related studies

Only a few studies have tried to assess the causal effect of sick leave on wages. The study most comparable to this paper is Hansen (2000), which exploits a policy change in Sweden as an instrumental variable for sick leave and finds substantial costs of sick leave for women but not for men. Hansen (2000) finds that, for women, each additional day of sick

[2] There is a large literature on *Compensating wage differentials* starting out from Rosen (1974). Wage compensation is often related to safety. A recent application, estimating workers marginal willingness to pay for safety can be found in Dale-Olsen (2005).

leave reduces the wage rate by 0.2 percent. He finds no effects on wages, however, of staying at home with a sick child, and interprets this as support for a signaling argument rather than having to do with human capital accumulation (Hansen, 2000, p.51). The approach of Hansen (2000) differs from this paper's approach in some important aspects. First, Hansen (2000) studies the effect of sick leave on current wages, implying that employers may not have had time to adjust wages. Second, the policy shift, which is the most credible of the instrumental variables used for identification, changes sick-leave rates for all workers from one year to another. The amount of information used to identify the causal effects is thus fairly limited. Finally, the dataset used is relatively small, which perhaps explains why Hansen (2000) only finds effects for female workers.

In a recently published paper Ichino and Moretti (2009) find that absences of women below the age of 45 tend to follow a 28 day cycle not present for older women or for men. They interpret this as absenteeism caused by the menstrual cycle, reflecting biological differences and not different propensities for taking occasional days off. From a model of statistical discrimination they hypothesize that the relationship between absence and earnings should be weaker for females than for males - because biologically caused cyclical absence makes sick leave a less informative signal of productivity and effort for females than for males - a proposition they find support for in their data. They estimate that one additional day of cyclical absence costs male workers about 2.5 percent in earnings, whereas the cost for female workers is 1.5 percent. Finally, they find that biological differences in cyclical absence can account for at least 14 percent of the gender wage gap.[3]

Differences in cyclical absences among workers should be interpreted as permanent – or at least long-lasting – worker heterogeneity. When workers with one additional day of

[3] Even if cyclical absenteeism is more costly for males than females, the "quantity effect" dominates the "price effect" as females have more cyclical absences than males.

cyclical absence earn 2.5 percent less, it is not the causal effect of one additional day of sick leave – but a manifestation of how heterogeneity in health (or effort) affects earnings over time. Ichino and Moretti (2009) convincingly illustrate the importance of biological differences and their importance in the labor market. However, their question is different from the question of this paper.

The rest of this paper is organized as follows. Section 2 discusses why sick leave is costly for individuals. Section 3 presents the data, the empirical specifications and the robustness tests. The results are presented in section 4, which also discusses some of the findings in detail. Section 5 concludes. The estimation procedure for the leniency indicator used as instrumental variable is described in more detail in the appendix.

2. Why should sick leave affect earnings and employment?

The relationship between sick leave and wages is investigated theoretically by - among others - Weiss (1985) and Coles and Treble (1993, 1996). In their models, firms whose absences are costly are willing to pay more - so that either their workers are less absent, or they are able to attract better workers. Examples are firms with assembly line production and firms where team production is important. Firms' sick-leave costs are studied by Nicholson et al. (2006) using a survey of 800 managers in 12 industries. They find that the cost "varies across jobs according to the ease with which a manager can find a perfect replacement for the absent worker, the extent to which the worker functions as part of a team, and the time sensitivity of the worker's output" (Nicholson et al., 2006, p.111). The study estimates the (non wage) cost of the median firm to be 28 percent of wages.

Sick leave hits employers financially and they may choose to punish the absentees.

One motivation can be to create incentives to work. Employers may link bonus arrangements or wage penalties to attendance. Another reason for punish absentees is to avoid attracting absence prone workers by building a reputation of having a low tolerance towards absence. We can imagine that some employers offer high effort/high pay contracts and that sickness absence is unpopular among their employees. Anecdotal “evidence” from e.g. law firms supports this. In some firms, employees work a lot, are very well paid and sickness absence is “suspiciously” rare. If high effort/high pay contracts are offered, we can also imagine that some employers offer low pay/low effort contracts. We know that sickness absence rates vary substantially between different types of work, and that sickness absence is negatively correlated with earnings.

A second, but related, way sick leave can affect earnings is through job promotion. Consider a firm pondering the promotion of a worker. A worker's sick-leave history seems likely to matter, since it discloses hidden information about the worker’s effort, productivity and health. If sick leave matters with regard to promotions, and a promotion in turn implies a wage rise, sick leave will affect future wages negatively.

A third possible relationship between sick leave and wages is through the process of specific human capital accumulation or depreciation. While one worker is away on sick leave, other workers may carry out important projects, making them more valuable to the firm than workers of otherwise equal value. In practice, however, it seems unlikely that being away from work for a few weeks affects human capital in such a way as to affect subsequent earnings.

Sickness absence is, almost mechanically, the main route out of the labor force. Nearly all workers who end up as disabled are initially on sick leave. Hence, the data show a strong correlation between sick leave and the probability of leaving the labor market. Such a correlation can have several origins – many of which are not causal in the sense that being on sick leave is the *reason* why workers leave employment and often end up as disabled. Arguably, the most obvious explanation for such a correlation is health. Workers become ill, they go on sick leave but do not recover and end up as disabled. Another explanation is that workers, who have a distaste for working, use sick leave as a gateway to disability pension. The first of these scenarios is the reason for having sickness insurance schemes. The latter is maybe why economists put so much effort into documenting moral hazard. But none of these scenarios apply any sort of causal interpretation to *being on sick leave*. Markussen et al. (2009) estimates the probability of returning to work after sick leave, and find strong negative duration dependence despite controlling for unobserved heterogeneity in a very flexible manner.[4] On the face of it, this suggests that being on sick leave has a negative causal impact on the probability of returning to work. However, this result depends critically on the proper functioning of the procedure for capturing unobserved heterogeneity – which is not easily tested. In order to distinguish an effect of sick leave from its underlying cause (health, preferences) it is necessary to exploit the discretionary nature of sickness absences and study situations where similar workers in similar situations show different sick-leave behavior. This is achieved by using physicians' leniency as an instrumental variable for individual sickness absence.

Sick leave may have a causal effect on subsequent employment for many reasons. The possibility of using sick leave to recover from an illness may enable a worker with health problems, otherwise unable to fill a normal job position, to remain employed. If empirically important, this means that sickness insurance increases employment. However, there may also

[4] Since health is unobserved, controlling for unobserved heterogeneity is crucial when estimating duration dependence. The workers who first return to work are the ones with best health, meaning that as duration increases, the remaining population on sick leave becomes more and more disadvantaged. This results, almost mechanically, in negative duration dependence.

be negative effects. Most explicitly, sick leave may increase the probability of being fired or laid off. This is studied by Hesselius (2007) who shows that workers with high sickness absence rates are more likely to become unemployed later on. The Hesselius study is not able, however, to separate the possible causal effects of sick leave from possible unobserved characteristics correlated with both productivity and sick leave. Supporting evidence is found by Henningsen and Hægeland (2008) who study mobility in downsizing firms and find that workers with a history of sick leave are more likely to leave. There may also be a less explicit mechanism at work. If employers respond by reducing wage growth and holding back on job promotions for workers on sick leave, this may reduce employees' work motivation. Employers may also prefer their absence prone workers leaving in order to be able to replace them with younger and healthier workers. Even if this is not explicit, there may be substantial heterogeneity in the amount of effort employers put into keeping absence prone workers employed.

There may also be negative effects on employment due to personal or psychological reasons. While absent, workers may develop *mental barriers* towards returning to employment. One possible reason could be a fear of distrust among their colleagues as to whether they are genuinely ill. Finally, sick leave may have a negative effect on employment due to habit formation. Workers may develop a "predilection for absence" – or a taste for domestic life.

3. Estimating the individual cost of sick leave

Sick leave, earnings and employment may be correlated for a number of reasons. In order to find causal effects of sick leave on earnings, holding individual factors such as motivation or health constant, we need an instrumental variable for sick leave. To be suitable, an instrumental variable should be correlated with sick leave but uncorrelated with the other unobservable variables that make sick leave an endogenous regressor when explaining earnings or employment. The instrumental variable proposed in this paper is the strictness or leniency of workers' primary care physicians, who provide medical certificates for all sick-leave spells lasting longer than 3 or 8 days. The estimation is carried out in three stages, described in detail below. In Stage 0, which is based on the findings of Markussen et al. (2009), a leniency indicator for Norwegian primary care physicians is estimated. In Stage 1, this indicator is used to explain sick leave and several robustness checks are discussed and conducted. Finally, in Stage 2, predicted values from Stage 1 are used to explain future earnings and employment.

3.1 Data

This paper makes use of Norwegian administrative register data. The data on sick leave includes start dates and end dates for all certified sick-leave spells in Norway from January 2001 through 2005, and are provided by the Social Security Administration (NAV). This dataset also includes the (encrypted) identity of the physician responsible for certifying each spell. The general rule is that sick leave lasting more than three days must be certified by a physician, although certification is not required until the 9th day for employees in firms participating in the so-called inclusive workplace agreement, which covers around half of the labor force. Note that self-reported sick-leave episodes, prior to the physician visit, are not observed by us.

Data on earnings, employment, education, sector of work, residency etc. stem from various sources but are all administrated and provided by Statistics Norway.

In this paper the population of interest is only the employed. In order to be included in the dataset one must be registered as a full-time employee and earn at least 1.85G[5] - approximately 16 000 USD – during at least one of the years 2001-2004. Furthermore, one must be aged between 25 and 59. The lower limit is set to handle the issue of young workers going from work to studies and back. The upper limit is set to avoid the age groups that qualify for early retirement programs – starting at age 62. To be defined as employed in year *t+2*, the same definitions are applied: one must be registered as a full-time employee and earn at least 1.85G. The data set is presented in Table 1.[6]

In total, the dataset contains 4.39 million employee/years, 60 percent of which have no registered sick-leave episodes in year *t*. Workers with certified sick leave are more predominantly females, and they are less educated than those with no sick-leave episodes. They also earn substantially less. Two years on, average nominal earnings for workers with sick leave in *t* are slightly reduced, while average earnings for those without sick leave have risen by 7 percent. The main reason for this is probably that just 82 percent of those with sick leave in *t* are still employed full-time (according to our definition) two years later, while the corresponding figure for those without sick leave is 90 percent. Whether or not these observations capture anything causal will be investigated below.

[5] G is a unit for calculation of social benefits in Norway that is adjusted for inflation.
[6] The dataset used by Markussen et al. (2009) to estimate the physicians' leniency indicator used as an instrumental variable is slightly different, the main differences being the inclusion of part-time workers and the exclusion of workers not aged between 30 and 60.

Table 1: Data descriptions

	All workers	Without sick leave in year *t*	With sick leave in year *t*
No. of worker/year obs.	4 388 269	2 610 461	1 777 808
Percent females	39.4	33.8	47.6
Age (mean)	42.5	42.8	42.1
Education			
< 10 years	6.2	5.6	7.1
> 13 years	38.0	42.3	31.8
Employed in *t+1*	90.0	92.3	86.5
Employed in *t+2*	86.8	90.3	81.7
Sick-leave rate	6.2	-	15.3
Earnings in *t*	385 465	410 816	348 241
Earnings in *t+2*	400 574	437 778	345 944

Notes: Included in the data sample are all workers employed full-time during one or more of the years *t=2001-2004* and earning more than 1.85 (approx: 16000 USD). Earnings in *t+1* and *t+2* are *unconditional* on employment in these years.

3.2 Estimating physician strictness and constructing the instrumental variable – Stage 0

Using an extraordinarily rich set of Norwegian administrative data for 2001-2005, Markussen et al. (2009) estimates a *leniency indicator* for Norwegian primary care physicians. In their paper, individual sickness absence propensity is modeled by means of a multivariate hazard rate model. In short, they explain individuals' sick leave by a rich set of individual factors, workplace fixed effects and physician fixed effects. These physician effects are transformed into a single measure, which is the leniency indicator used as instrumental variable in this paper. Since this paper makes direct use of their estimates, a detailed description of their estimation strategy is provided in the appendix.

The instrumental variable, which will be denoted by z, is available for 3205 physicians and there is substantial variation in physicians' leniency, causing variations in the predicted sick-leave rates of a worker - conditional on physician only - from below 4 and up to 14 percent. However, most workers are registered with physicians with expected sick-leave rates of 6 to 10 percent. The variation is illustrated in the upper panel (a) of Figure 1. Panel (b) shows the distribution of individual sick-leave rates for 2003. The maximum value, 100 percent, implies that the worker is continuously on sick leave for an entire year. Most workers have no sick leave and the distribution is very different from the instrumental variable *z*.

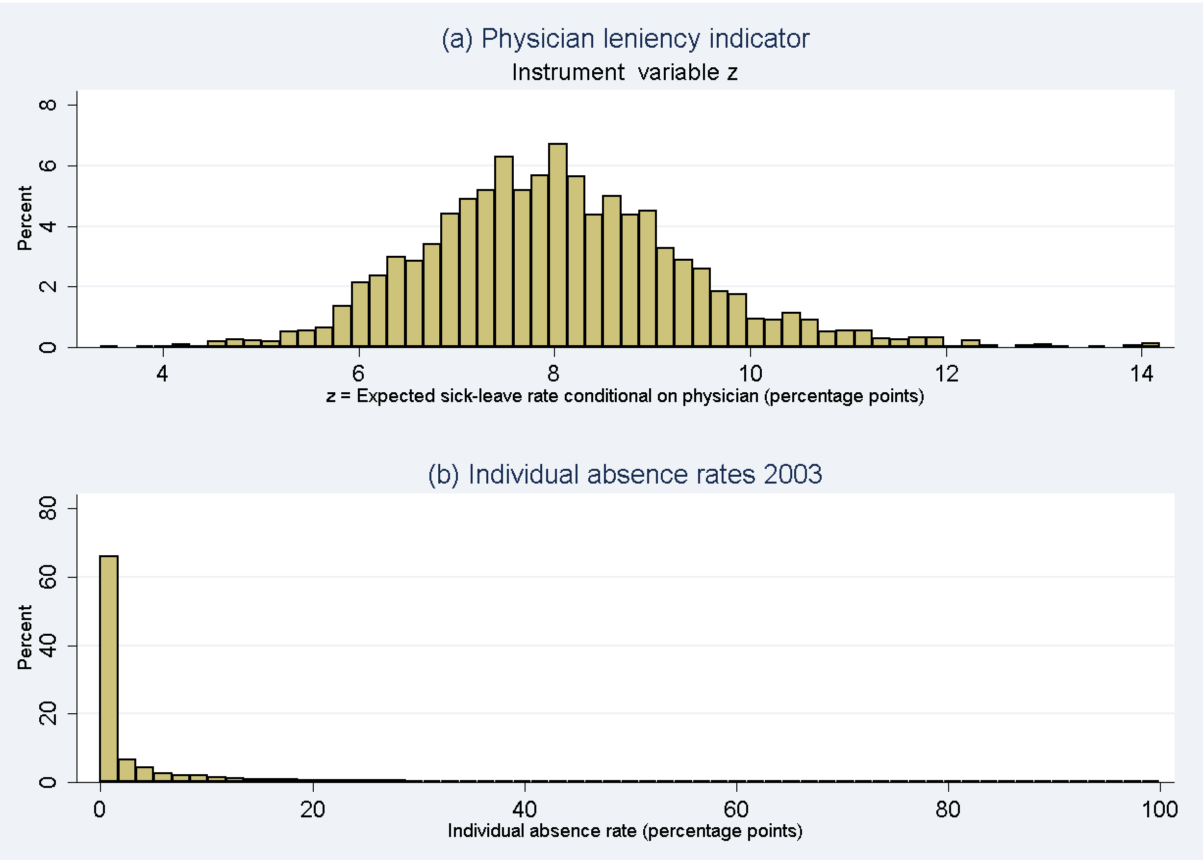

Figure 1: Distribution of the instrumental variable "physician leniency" and the instrumented variable "the individual sick-leave rate". Panel (a) displays the instrument variable *physician leniency (z)* estimated by Markussen et al. (2009). This variable should be interpreted as the expected sick-leave rate of a worker conditional on nothing but his physician. Panel (b) displays the distribution of individual sickness absence rates in 2003 for the workers in the sample. The distribution is highly skewed as most workers have no sick leave.

Substantial variation between physicians in their sick-listing practices is also found in medical studies, typically studying artificial case vignettes which is tested on different physicians (see e.g. Englund et al., 2000; Gorter et al., 2001).

Leniency or quality?

An alternative interpretation of the leniency indicator is that differences in sick leave across patients stem from differences in the *quality* of the medical services provided. To investigate the relationship between physician leniency and quality, I follow Biørn and Godager (2009) and estimate age and gender adjusted mortality rates for each physician's patients and use these as a proxy for physician quality. To be more precise, I construct a dataset for all Norwegian citizens above 40 years of age in the years 2001-2005 connected to one of the physicians for whom the leniency indicator is available. On this dataset I fit a linear probability model where the endogenous variable is an indicator taking 1 if the person dies during the relevant year and zero otherwise. I explain individual mortality with gender (1 dummy), age (69 dummy variables), year (4 dummy variables) and physician (3205 physicians). These physician coefficients are then a proxy for physicians' quality. Figure A1 plots physician quality against leniency together with a regression line which is not significantly different from zero. It turns out that physicians' leniency and quality, measured this way, are not particularly related, with a coefficient of correlation of 0.0125. I have also tested whether including physicians' quality among the covariates in the models to be estimated in any way affects the results, which it does not.[7]

The use of generated explanatory variables

An econometric issue worth discussing is the use of *generated explanatory variables*. Using the notation from Wooldridge (2002, p.117) we have an instrumental variable $z = g(\mathbf{w}, \lambda)$ where *g(.),* is a known function, **w** is a vector of observed variables and λ is unobserved and needs to be estimated. The generated instrument is then $\hat{z} \equiv g(\mathbf{w}, \hat{\lambda})$. "[U]nder

[7] As is to be expected, physicians' quality is, statistically, significantly (negatively) related to individual sick leave. However, including quality among the regressors, or as an additional instrumental variable, has no impact on the results to be presented.

conditions that are met in many applications, we can ignore the fact that the instruments were estimated (…) for inference." When $E(u \mid \mathbf{w}) = 0$, where u is the error term in the second stage equation, "the $\sqrt{N}$ -asymptotic distribution of $\hat{\beta}$ is the *same* whether we use λ or $\hat{\lambda}$ in constructing the instruments. This fact greatly simplifies calculation of asymptotic standard errors and test statistics." (Wooldridge, 2002, p.117).

3.3 Explaining individual sick-leave behavior by physician strictness – Stage 1

The first-stage equation to be estimated is given by (1.1) where a_{it} is the sick-leave rate of worker i in year t, y_{it} is earnings the previous year, c_{it} is a set of individual control variables such as age, gender, education, sector or employment, calendar time etc. and z_{it} is the instrumental variable, physicians' leniency, as described above.

$$a_{it} = y_{it-1}\delta + c_{it}\beta + z_{it}\gamma + e_{it} \tag{1.1}$$

When Markussen et al. (2009) estimate the coefficients used to construct the instrumental variable *z,* each physician's leniency is measured on the basis of the sick-leave behavior of their patients. Unfortunately, visits by a worker to a physician to ask for a medical certificate are not observed in their study. They only observe the outcome when sick leave was certified. Hence, after controlling for observable characteristics we have to assume that differences in (mean) sick leave of each physician's patients can be attributed to differences in the physicians' leniency, and not to unobserved characteristics, as far as these patients are concerned. This assumption, which is a strong one, is the core of the *identifying strategy* in this paper (Angrist and Pische, 2009, p.7). One should keep in mind, however, that the set of observable characteristics is rich and, in particular, that it includes workplace fixed effects.

A series of robustness tests are carried out in order to test whether the identifying assumption is seriously violated. In particular, three possible problems are examined; (i) geographical confounding factors that may be picked up by the physician effects, (ii) endogenous matching of physicians and workers, and (iii) various types of what is often referred to as the *reflection problem* (Manski, 1993).

(i) Geographical confounding factors

A potential problem is that geographical confounding factors may have been picked up by the physician fixed effects. One example of this problem occurs in large cities. It is well known that there are substantial differences in life expectancy within Norway's capital, Oslo. These differences are related to observable characteristics, such as income and education, but they are also related to unobservable characteristics such as lifestyle. If people with, say, a "taste for exercise" tend to live near the recreational areas while those with a "taste for bars and nightlife" tend to live in the city centre we encounter a problem if people also tend to have a physician who is located in their neighborhood. These workers, with different preferences and lifestyles, may very well work at the same place and be similar along all other observable lines. To control for this, a set of "census tracts" (grunnkretser), used for Norwegian regional statistics, are included. The workers in the dataset used for estimation live in around 13 500 such areas and the model is estimated with *census tract fixed effects.* One should keep in mind that Norway has a mere 4.5 million inhabitants, so these census tracts are quite small.[8]

[8] The workforce consists of roughly 2 million workers, hence the average number of workers in each census tract is around 150.

(ii) Endogenous matching of patients and physicians

Another potential problem with using physicians' leniency as an instrumental variable is that there might be a sorting of employees into physicians. Using a simple Internet service, Norwegian citizens can change their primary care physician provided that the new physician has free capacity. This may very well lead to a situation where poorly motivated workers choose the most lenient physicians, and if this is empirically important it makes the instrumental variable unsuitable. The importance of this sorting was studied by Markussen et al. (2009). Using a variance decomposition exercise, they find no evidence of sorting into physicians along *observable* variables. Close to none of the variation in sick leave among physicians was related to differences in education, workplace, income etc. We should expect unobservable variables such as motivation and health to be correlated with individual characteristics such that the result, that these observables are unimportant for explaining variation between physicians, is encouraging. However, to test further for such sorting the model is estimated on a subsample consisting only of workers who were collectively transferred to a new GP such that these workers did not choose their physician themselves. Such situations occur when a GP retires, moves to another part of the country or for some other reason decides to quit his practice.

(iii) The reflection problem

A third potential problem is often referred to as the *reflection problem* (Manski, 1993). In essence, the instrumental variable used is the mean absence rate of each physician's patients, adjusted for a number of observed characteristics. Hence, each individual's absence rate is used to construct the instrumental variable that is meant to be exogenous to individual unobserved factors. In principal, this clearly violates the conditions for a valid instrumental variable, but we should expect the problem to be less serious as the groups become large. A related problem is that variation in the instrumental variable may just reflect statistical randomness or "noise". Sick leave is a stochastic process and even if physicians had no impact at all, the instrumental variable would not be zero since it would capture some of the variation not captured by other variables. Markussen et al. (2009) test for this using a placebo regression: Workers are divided into groups of the same size as the patient populations belonging to each physician and the model is re-estimated. If the physician fixed effects capture mostly randomness, the variation of the placebo coefficients should have been roughly the same as for the *real* coefficients. They find that most of the variation between physicians seems to be caused by the physician rather than by randomness. Moving from the 10th to the 90th percentile in the physicians' leniency distribution raises the representative employee's absence rate from 4.5 to 7.1 percent; i.e., by 58 percent. In the placebo estimation, the difference is only 17 percent. Hence, again, most of the variation across physicians seems to be causal in the sense that it is related to the physicians. To test whether *reflection* and *randomness* are driving the results, a third robustness exercise is carried out, referred to as the *split sample approach*. First, each physician's patient lists are randomly split into two parts. The first group is used to estimate a new physician indicator using a simple linear regression model, where a_{it} is the sick-leave rate of worker i in year t, y_{it} is earnings the previous year, c_{it} is a set of individual control variables such as age, gender, education, sector or employment, calendar time etc. and P_{it} is a set of dummy variables, one for each physician.

$$a_{it} = y_{it-1}\delta + c_{it}\beta + P_{it}\eta + e_{it} \quad \forall i \in Group1$$

The set of coefficients $\boldsymbol{\eta} = \{\eta_1, ..., \eta_P\}$ is a physician leniency indicator estimated from half the dataset, and the first step equation (1.1) is estimated using $\boldsymbol{\eta} = \{\eta_1, ..., \eta_P\}$ as an instrumental variable on the half of the dataset that was not used when the strictness indicator was estimated (i.e. *Group 2*).

Estimation results from the first stage

The results from estimating the first stage is presented in Table 2 together with several robustness checks in line with the preceding discussion.

Table2

Estimation of Stage 1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Physician strictness (z_{it})	0.836	0.813	0.767	0.750	0.752	0.722	0.591
	(0.017)	(0.015)	(0.013)	(0.016)	(0.011)	(0.016)	(0.112)
Past earnings (y_{it-1}),	-	-0.008	-0.004	-0.005	-0.004	-0.004	-0.005
(kroner x 1000)		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Individual characteristics			Yes	Yes	Yes	Yes	Yes
Large firms only				Yes			
Census tract fixed effects (13424)					Yes		
Split sample						Yes	
Only exogenous physician switches							Yes
R-squared	0.006	0.017	0.039	0.052	0.048	0.039	0.035
No. observations	4388222	4388222	4388222	1965814	4387026	2192975	21036

Notes: Table 2 displays the results from estimation of (1.1). Column 3 is the baseline specification. Columns 1 and 2 are not controlled for individual characteristics. Columns 4 – 7 are robustness tests described in the text above. Robust standard errors clustered on physician are reported in brackets. Individual characteristics are age (linear, squared and cubed), years of education (linear, squared, cubed), low skill and high skill (dummy variables), marital status (3 dummies), gender, country of origin (2 dummies), sector of employment (11 dummies), calendar year (3 dummies)

Column (3) displays the baseline estimates of (1.1) where individual characteristics and past earnings are included as control variables. The coefficient on the instrumental variable z_{it} is 0.767. When Markussen et al. (2009) estimated z_{it} they used workplace fixed effects for all workplaces with at least 100 employees. For smaller firms they used controls for sector and size. A substantial fraction of Norwegian employees work for relatively small firms such that for many workers, the instrument z_{it} is estimated without controlling for workplace fixed effects. In column (4) the sample is thus restricted to only include workers in firms large enough to be represented by a separate dummy variable when estimating z_{it}. The sample is reduced by around 50 %, but the estimated coefficient is not significantly different from the one in column (3) where all workers were included. When controlling for 13 424 geographical areas (census tracts), as shown in column (5), the estimate is nearly unchanged suggesting that the instrumental variable z_{it} is not just capturing geographical variations of some kind. Column (6) displays the first stage estimates using the split sample approach. Despite the coefficient being fairly similar to the other columns, the first stage estimates are not directly comparable as the instrumental variable is different.[9] The second stage estimates are comparable, however, and are presented in Table 3. Finally, column (7) displays the results from a substantially smaller subsample consisting only of workers whose physician is "new" and for whom this change occurred for an exogenous reason. The sample is reduced from 4.39 million to 21 036 worker/year observations and from 3205 to 124 physicians, such that the standard errors are larger. The estimated coefficient for z_{it}, however, is not statistically different from the others.

[9] When the first stage equation is estimated using "*Group 1*", the group used to estimate the split sample leniency indicator, the coefficient is approximately 1 – which is to be expected.

3.4 Explaining future earnings by current sick leave – Stage 2

In the second stage, predicted values for a_{it} are used as regressors, together with a set of individual characteristics x_{it} to explain earnings or employment two years later, referred to as y^k_{it+2}, where *k*=***Emp****loyment,* ***E****arnings*. The model estimated in Stage 2 is given by (1.2).

$$y^k_{it+2} = \delta_1 y^E_{it-1} + x_{it}\delta_2 + \beta \hat{a}_{it} + \varepsilon_{it+2} \tag{1.2}$$

Several questions arise regarding how the 2^{nd} stage equation should be specified. Below four such issues are discussed: (i) inclusion of individual fixed effects, (ii) the non-standard dynamic specification, (iii) choice of functional form, and (iv) inclusion of control variables.

(i) Should individual fixed effects be included?

The model is not estimated with individual fixed effects, for two reasons. First, fixed effects would implicitly exclude all workers not present in the data for more than one year. Such exclusion would create a potentially serious sorting problem, as we condition on future employment – which is an endogenous variable. For estimating effects on future employment this is of course meaningless. Second, individual fixed effects would also exclude all workers who did not change physicians during the observation window. The reason Markussen et al. (2009) report that sorting into physicians is not empirically important (along observable characteristics) may be that most workers do not change physicians very often. Hence, by including only the subset of workers that do change physicians, the problem with sorting of employees into physicians may become much more serious than otherwise. To still capture as much individual heterogeneity as possible, past earnings are included in the model, denoted y^E_{it-1}.

(ii) The non-standard dynamic specification

There are good reasons for specifying the wage equation in such a non-standard manner. Normally wage changes occur through annual wage negotiations. Hence, in many cases sick leave in year *t* will not affect wages in year *t* at all if the episode occurred after wages had been negotiated.[10] Even if negotiations were affected, such wage changes are usually phased in, such that there is a substantial lag before the full change is implemented. Finally, since we observe earnings annually, the full effect is not observed until the full wage change has been implemented for an entire year, which is in *t+2*. The intuition behind this reasoning is also confirmed in the data. When estimating variants of (1.2) where the dependent variable is replaced by y_{it}^{k} or y_{it+1}^{k}, the effects are considerably smaller than what is displayed in Table 3, in line with the reasoning above.[11]

To study effects on employment we must study the effects on employment at a later stage since all workers are employed in year *t* by definition. Since workers can remain on sick leave for up to a year, and employment is registered annually, the first year it seems reasonable to study employment effects of being on sick leave in year *t*, is in year *t+2.*

(iii) Choice of functional form

To study the effects of sick leave on subsequent earnings the dependent variable must be specified in somewhat more detail. The arguably most common approach is to use the logarithm of earnings as the dependent variable. Income distributions are often approximately log-normally distributed such that the distribution of the logarithm of earnings is

[10] In year *t* there is also another potential effect on earnings from reduced overtime payment and lost bonuses for absent workers, which is not within the scope of this paper.

[11] The effects using the baseline specification of (1.1) and (1.2) displayed in column 3 in Table (2) and Table (3), only changing the time specification of the dependent variable is: y_{it}: $\beta = -639.3$, y_{it+1}: $\beta = -1517.7$, y_{it+2}: $\beta = -2270.4$, when the model is estimated on earnings in NOK, and: y_{it}: $\beta = -0.0015$, y_{it+1}: $\beta = -0.008$, y_{it+2}: $\beta = -0.012$, when the model is estimated on log earnings.

approximately normal – making it suitable for linear regression. However, the data sample is constructed on the basis of employment in year *t*. In year *t+2* some of the workers are no longer employed, neither full-time nor part-time, and have zero earnings. Using the logarithmic transformation means that workers with zero earnings are excluded from the estimation. This creates a potential sorting problem as we, implicitly, condition on endogenous variables. To be more specific, if the probability of having zero earnings in *t+2* increases for the predicted sick-leave rate $\hat{a}_{it}$, the workers with high $\hat{a}_{it}$ for whom we observe earnings, are positively sorted in the sense that that they have drawn a high error term ε_{it+2}^{k}. In practice, a fairly limited fraction of workers exits the dataset when log earnings are computed. The number of observations is reduced from 4.4 to 4.3 million. A different approach, which can be thought of as the *experimental* approach, is simply to include earnings linearly such that zero earnings are included. Then the sorting problem is eliminated. However, in this case the model is in a sense misspecified as the dependent variable ends up with a peculiar distribution. Results from both approaches will be presented below.

To estimate the effects on employment in *t+2*, a linear probability model (LPM) is used. The dependent variable y_{it+2}^{Emp} takes two values; $y_{it+2}^{Emp} = 1$ if the worker is also employed full-time in *t+2*, and $y_{it+2}^{Emp} = 0$ otherwise.

(iv) The choice of control variables

Earnings in *t* and *t+1* are excluded as they are partly a function of a_{it}. As a model for earnings, an regular AR-model where these variables were included would most certainly do a better job than the model suggested here. However, the specification is chosen to construct a quasi natural experiment setting where otherwise equal workers are observed at a later stage –

conditional on nothing but their predicted sick-leave rate. Sick leave in *t+1* is also excluded, as it may be affected by the interaction between sick leave, wage negotiations and earnings in year *t*. The remaining individual characteristics are also observed in *t*.

In order to investigate whether the effects of sick leave on earnings and employment differs between sectors and for different groups of workers, estimates for separate subgroups are also presented. These groups are then, with no exceptions, made conditional on the workers' group status in year *t* in order to avoid the problem of conditioning on endogenous outcomes.

4. Results

Table 3 presents the main results along with the same series of robustness checks as presented above for the first stage estimates. In panel (A) the dependent variable is earnings in *t+2* measured in kroner, whereas in panel (B) the dependent variable is the logarithm of earnings in *t+2*.[12] In panel (C) the dependent variable is a binary variable taking 1 if the worker is employed full-time in *t+2* and zero otherwise. Predicted sick leave is measured in percent (from 0 to 100) such that its coefficients should be interpreted as the change in the dependent variable if a worker's sick-leave rate increases with 1 percentage point.

[12] In the first stage, equation (1.1) is estimated as displayed in Table 2, but in the model with log earnings as the dependent variable, log earnings (t-1) are used as control.

Table 3

Estimation of Stage 2

(A) Dependent variable: earnings in t+2 in kroner

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted sick leave (a_{it})	-5937.2	-3111.2	-2270.4	-2629.6	-2004.8	-2744.6	-2273.0
	(823.3)	(242.1)	(219.8)	(291.6)	(174.5)[13]	(239.5)	(2014.0)
Past earnings (y_{it-1}), kr		0.841	0.785	0.801	0.773	0.783	0.719
		(0.010)	(0.011)	(0.016)	(0.011)	(0.015)	(0.037)
Individual characteristics			Yes	Yes	Yes	Yes	Yes
Large firms only				Yes			
Census tract fixed effects (13424)					Yes		
Split sample						Yes	
Only exog. physician switches							Yes
R-squared	0.018	0.479	0.498	0.541	0.491	0.496	0.530
No. observations	4388222	4388222	4388222	1965814	4387026	2192975	21036

(B) Dependent variable: logarithm of earnings in t+2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted sick leave (a_{it})	-0.020	-0.015	-0.012	-0.013	-0.011	-0.013	-0.012
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.009)
Past earnings (y_{it-1}), log		0.579	0.502	0.486	0.487	0.500	0.517
		(0.004)	(0.003)	(0.004)	(0.001)	(0.003)	(0.030)
Individual characteristics			Yes	Yes	Yes	Yes	Yes
Large firms only				Yes			
Census tract fixed effects (13424)					Yes		
Split sample						Yes	
Only exog. physician switches							Yes
R-squared	0.091	0.264	0.291	0.322	0.240	0.292	0.298
No. observations	4298134	4276802	4276802	1917250	4276788	2137433	20417

[13] The models in column (5), in all three panels, are estimated using the command *areg* in STATA, and the 1st and 2nd stages are estimated separately, indicating that the reported standard errors are incorrect.

(C) Dependent variable: full-time employment in t+2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted sick leave (a_{it})	-0.006	-0.005	-0.005	-0.004	-0.005	-0.004	-0.009
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)
Past earnings (y_{it-1}), kr		1.6e-7	1.2e-7	9.8e-8	1.2e-7	1.22e-7	1.4e-7
		(5.6e-9)	(4.2e-9)	(4.8e-9)	(4.4e-9)	(5.2e-9)	(3.5e-8)
Individual characteristics			Yes	Yes	Yes	Yes	Yes
Large firms only				Yes			
Census tract fixed effects (13424)					Yes		
Split sample						Yes	
Only exog. physician switches							Yes
R-squared	0.055	0.063	0.077	0.082	0.042	0.075	0.059
No. observations	4388222	4388222	4388222	1965814	4387026	2192975	21036

Notes: Table 3 displays the results from estimation of (1.2). The corresponding *first stage estimates* are displayed in Table 2. Column 3 is the baseline specification. Columns 1 and 2 are not controlled for individual characteristics. Columns 4 – 7 are robustness tests described in Section 3.3 above. Robust standard errors clustered on physician are reported in brackets, Individual characteristics are age (linear, squared and cubed), years of education (linear, squared, cubed), low skill and high skill (dummy variables), marital status (3 dummies), gender, country of origin (2 dummies), sector of employment (11 dummies), calendar year (3 dummies).

Column (3) shows the preferred specification of the model and columns (4)-(7) show the different robustness tests. On average, a one percentage point increase in sick leave reduces subsequent earnings by 2271 NOK, approximately 405 USD. Using the log specification in Panel (B), the same cost is estimated to 1.2 percent. The estimated effects are robust to the specification using census tract fixed effects (13 424 groups), the estimation strategy where the sample is split in two to avoid the reflection problem, and the subsample consisting only of workers whose new physician was a result of a physician-induced change in the patient-physician relation.[14] The estimate in this subsample (column 7) is, however, not

[14] The effects are also robust to combinations of these, such as the split sample approach *and* census tract fixed effects, or the split sample approach *and* physician-induced switches. To investigate the empirical importance of the reflection problem in this application, the model using the split sample physician leniency indicator is also

statistically different from zero. Note that the adjustment of the standard errors is particularly strict in this case, however, as the number of physicians is just 124 (109 in the log model).

Finally, Panel (C) shows that a part of this cost is due to the probability of ending employment (full-time) increasing by 0.5 percent. Hence, these estimates confirm the estimates of Markussen et al. (2009) of sick leave having a negative causal effect on future employment – using a completely different identification strategy.[15]

In an attempt to separate the effects on earnings from the effects on employment, the model is also estimated on a subsample made conditional on employment in *t+2,* such that the sample consists only of workers where $y_{it+2}^{Emp} = 1$. Since this conditioning is an endogenous variable, this is econometrically unsound and may bias the estimated costs towards zero as the group of workers that remain employed despite high sick leave is a selected sample. Nevertheless, the estimated costs in this group, using estimator (3) in Table 3, is -1411.6 NOK and -0.0029 log points. Both estimates are precisely estimated. This indicates that the employment effects account for around half the estimated earnings costs.

Economic significance

Interestingly, sick leave seems to have a causal effect both on earnings and employment, and these effects are substantial. To illustrate their economic significance, consider the following example: Earnings are often considered a persistent autocorrelated process with a drift. If so, a reduction in earnings one year will leave a scar for future periods. Consequently, a seemingly small reduction in earnings two years after sickness may accumulate to a large amount over time. If one assumes an AR1 process for earnings, and that

estimated on the same group used to estimate this indicator. The estimated coefficients from the second stage are very similar in both groups, indicating that this problem has limited empirical relevance in this application.

[15] There is a potential caveat regarding the results on employment. If lenient physicians are also lenient with respect to certifying disability insurance, this may lead to the same results, but with a different interpretation. There are, however, strong arguments why this should be a minor problem. First, it typically takes many years before a worker becomes disabled (more than two). Second, qualifying for disability pension, also involves a "neutral" physician, employed by the social security administration.

interest rates, earnings growth, and discounting cancel each other out, the present value of such a loss of earnings, measured as the share of current earnings, can be written as in (1.3). The *gross* cost *C* is a function of the remaining *T* years on the labor market, the estimated earnings reduction from sick leave β (the log model) and the coefficient of autocorrelation ρ.

$$C(T,\beta,\rho)=\beta\sum_{t=1}^{T}\rho^{t-1} \tag{1.3}$$

In Table 4, these gross costs of sick leave are illustrated for workers with various numbers of years left on the labor market and for different levels of ρ.

Table 4

Total cost of a one percentage point increase in sick leave for different degrees of persistence in earnings

Working years left	ρ = 0.95	ρ = 0.9	ρ = 0.8	ρ = 0.5	ρ = 0.0
1	0.01	0.01	0.01	0.01	0.01
10	0.09	0.07	0.05	0.02	0.01
20	0.14	0.10	0.05	0.02	0.01
30	0.17	0.11	0.05	0.02	0.01

Note: The gross marginal cost of sick leave, using (1.3) and $\beta=.011$, measured as the share of earnings in *t+2*.

For a worker earning USD 60.000 a year, with 10 years left on the labor market, a one percentage point increase in sickness absence will cost from \$6 600to \$600,depending on the persistence of earnings. Without any sickness insurance payments, the loss of current earnings from such an increase in sickness absence would be around \$420after taxes.[16] Hence, the reduction in future earnings more than weighs up for the lack of incentives in the generous

[16]\$60 000 divided by 220 working days, excl. 30 percent tax.

sickness insurance arrangements in Norway.

Non-linear effects?

Is the marginal effect of sickness absence linear in spell duration? A signaling argument would suggest the opposite. If we consider sick leave a measure of effort, zero sick leave is the maximum effort one can provide. Hence, as a measure of effort sick leave is censored. The marginal cost of going from zero to a positive amount of sick leave may thus be higher than when the number of sick leave days is raised from an already positive level, as the worker no longer is associated with the "max-effort/no sick leave" pool of workers. Following Angrist and Pishcke (2009, p.192) the model is extended with an additional first stage equation where the instrumental variable *z* is squared. The model is then as given by (1.4) – (1.6).

$$y_{it+2}^{k} = \delta_1 y_{it-1}^{E} + x_{it}\delta_2 + \beta\hat{a}_{it} + \beta\hat{a}_{it}^{2} + \varepsilon_{it+2} \tag{1.4}$$

$$a_{it} = \delta_1 y_{it-1}^{E} + z_{it}\gamma + x_{it}\delta_2 + u_{it} \tag{1.5}$$

$$a_{it}^{2} = \phi_1 y_{it-1}^{E} + z_{it}^{2}\varphi + x_{it}\phi_2 + e_{it} \tag{1.6}$$

The results are displayed in Table 5 and suggest that the effect is strongly non-linear and that the marginal cost of sick leave decreases with time. One should keep in mind that the instrumental variable has a fairly limited range as displayed in Figure 1. Hence, extrapolating the estimated polynomial to the limits of sick-leave rates equal to zero or 100 gives no meaning.

Table5

Non-linear effects			
	Linear	*Log*	*Employment*
Predicted sick leave (a_{it})	-2878.4	-0.208	-0.074
	(3737.8)	(0.015)	(0.007)
Predicted sick leave	11.3	0.004	0.0013
squared $\left(a_{it}^2\right)$	(73.1)	(0.000)	(0.0001)
No. observations	4388222	4276802	4388222

Notes: Table 5 displays the results from estimation of (1.4)-(1.6) allowing for a non-linear relationship between sick leave and earnings/employment. Standard errors (not clustered on physician) are reported in brackets, individual characteristics are age (linear, squared and cubed), years of education (linear, squared, cubed), low skill and high skill (dummy variables), marital status (3 dummies), gender, country of origin (2 dummies), sector of employment (11 dummies), calendar year (3 dummies).

Heterogeneous effects

The size of this dataset enables us to divide the data into subgroups and investigate whether these effects are different for different groups of workers. In the limited literature available in this field, conflicting results are reported regarding gender differences. Hansen (2000) reports that while females' wages are reduced following sick leave, male workers' wages are not affected. In contrast, Ichino and Moretti (2009) find that cyclical sick leave is less costly for females than for males. Hence, it seems natural to investigate whether these costs differ for men and women. From Table 4 it is evident that the extent to which these costs are substantial depends on the number of years workers have left in the labor market. Even though sick leave can be less costly for young workers two years later they can still be disciplined, whereas wage growth is less important for older workers. In order to learn more about how these effects differ between workers, the model is also estimated on subgroups grouped by education, earnings and jobs.[17] These results are displayed in Table 6.

In line with the results from Ichino and Moretti (2009), sick leave is substantially more expensive for males than for females, regardless of whether we measure earnings in kroner or logs. A one percentage point increase in sick leave for males leads to 1.5 percent lower earnings. For females, the same increase in sick leave reduces earnings by 0.9 percent. Since males tend to earn more than females, the relative difference using the linear estimator is even greater. Interestingly, the effect on employment is similar for men and women and cannot explain the gender difference. Table 7 and the following discussion below further explores these gender differences.

The pro anno cost of sick leave decreases with age. This fits nicely with the idea that the cost is permanent and young workers are disciplined by a smaller cost per year because they have more time left in the labor market. However, the estimates for employment also indicate that reduced employment can explain at least part of this difference. Young workers' (<30 years) employment is not affected by sick leave whereas sick leave seems more and more important for understanding withdrawal from the labor market as workers get older.

[17] Unfortunately, we do not observe ”jobs” in the data. The jobs described in Table 6 are defined using a combination of sector of employment and education.

Table 6

Different effects for different workers

		No. obs.	*Linear model*	*Log model*	*Employment*
Gender	Males	2661060	-3115.2***	-0.015***	-0.005***
	Females	1727162	-1356.1***	-0.009***	-0.005***
Age	< 30	276739	-1316.6**	-0.004	-0.001
	30-40	1545476	-1832.7***	-0.010***	-0.004***
	40-50	1393235	-2404.1***	-0.012***	-0.005***
	50-60	1172772	-2530.9***	-0.014***	-0.006***
Education	< High school	1485615	-2033.3***	-0.014***	-0.005***
	High school	1234308	-1889.6***	-0.011***	-0.005***
	Some college	1273391	-2465.1***	-0.009***	-0.003***
	Full college	394908	-8534.8***	-0.023***	-0.003
Earnings in *t*	Quartile 1	1097053	-1370.4***	-0.012***	-0.006***
	Quartile 2	1097059	-1505.5***	-0.010***	-0.005***
	Quartile 3	1097049	-1565.0***	-0.010***	-0.004***
	Quartile 4	1097061	-6546.7***	-0.014***	-0.002***
Jobs[17]	Teacher	356505	-2544.5***	-0.009***	-0.004***
	Nurse assistant	291491	-1289.4***	-0.008***	-0.005***
	Nurse	256993	-627.4***	-0.005***	-0.003***
	Physician	47940	-4783.5*	-0.021***	-0.003
	Civil servant	205027	-4046.8***	-0.014***	-0.003**
	Production/industry, low skill	554409	-2349.0***	-0.015***	-0.005***
	Production/industry, high skill	132497	-5898.4***	-0.015***	-0.003

Jobs (cont.)	R&D, high education	25209	-5677.8**	-0.022**	-0.000
	Trade & Commerce, low skill	534917	-1747.9***	-0.014***	-0.006***
	Trade & Commerce, high skill	114247	-2118.9	-0.019***	-0.003
	Banking & Finance, low skill	72750	-3519.9**	-0.009***	-0.004**
	Banking & Finance, high skill	57195	-14989.5**	-0.016***	-0.002

Notes: Table 6 displays the results from estimation of (1.2) for various subgroups of workers. All results stem from the baseline specification of the model, reported in Column 3 in Tables 2 and 3. (The first stage equation (1.1) is also estimated separately for each sub-group but the results are not reported). The numbers of observations are reported for the linear model and are slightly lower for the log specification because workers with zero earnings are excluded. Robust standard errors clustered on physician are reported in brackets. Individual characteristics are age (linear, squared and cubed), years of education (linear, squared, cubed), low skill and high skill (dummy variables), marital status (3 dummies), gender, country of origin (2 dummies), sector of employment (11 dummies), calendar year (3 dummies).

The estimates arising from separating the sample after education or earnings show a roughly similar picture. Earnings are most sensitive to sick leave for workers in either end of the skill distribution or earnings distribution – but for different reasons. At the lower end employment is more sensitive to sick leave, and this can explain, at least partly, why these workers' earnings are more sensitive to sick leave than others. At the upper end employment is affected much less indicating that wages – not hours worked – are affected.

There are large differences across "jobs". A nurse's earnings are only reduced by 627 kroner, or 0.5 percent, from a one percentage increase in sick leave, while the corresponding figure for a physician is 4784 kroner (2.1 percent). Overall, earnings in public sector jobs, such as teacher, nurse assistant, nurse or civil servant, are generally less affected than earnings in private sector jobs, such as production, R&D, Trade & Commerce and Banking & Finance. These differences do not seem to be driven by differences in effects on employment, which are roughly the same or larger in the public sector. This indicates that sick leave is taken into count when wages are set and promotions are given – and more so in the private sector where employers are more at liberty to take such considerations into account.

The gender puzzle

A striking observation is that sick leave seems to affect earnings for female workers far less than for male workers. We have also seen that sick leave is more costly for high earners and private sector employees. An obvious question is then whether the reason why sick leave is more costly for males is just because men are overrepresented among high-earners in the private sector. Such a hypothesis can easily be tested by estimating the model on subsamples where we compare females and males belonging to the same group. The results from such a strategy are displayed in Table 7, showing several interesting patterns. First, in typically female dominated jobs, such as teachers, assistant nurses and nurses, females' earnings are much less affected by sick leave than males' earnings. At the same time, females' employment is affected about as much as males'. Among civil servants, where we have a more even distribution of male and female workers, we see roughly the same pattern. But, in typically male dominated jobs, such as production and within trade & commerce, things are different. Here, females' and males' earnings are affected more or less equally, and the negative effect on females' employment is much stronger for the female dominated jobs and also stronger than the employment effects for men. A possible explanation for this pattern is that some jobs are "specialized" in having female workers. Females generally have substantially higher sick-leave rates than males, and this seems to be more accepted in some jobs than in others. In these jobs females' earnings are less affected, but so is employment. In male dominated jobs, females are "treated as males" and part of the outcome is reduced employment.

Another potential explanation as to why sick leave is less costly for females is that females' absences are more tolerated because they are partly related to pregnancies and taking care of small children. For these reasons sick leave may be less of a signal of low effort for females, and should therefore not be punished the same way as for males. This is also argued by Hansen (2000) who finds that own sickness reduces wages, but not sick leave due to sick children. Unfortunately, the data used in this paper do not cover sick leave due to sick children.[18] To investigate whether children can be the reason why sick leave is less costly for females, the models are estimated separately on females below 40 years of age, with and without children.

Table 7

The gender puzzle

		No. obs.	*Linear model*	*Log model*	*Employment*
Teacher	Males	150868	-3931.4***	-0.013***	-0.005*
	Females	205637	-1943.9***	-0.008***	-0.004***
Nurse assistant	Males	65241	-2898.2***	-0.013***	-0.004***
	Females	226250	-881.2***	-0.007***	-0.006***
Nurse	Males	47756	-1814.2**	-0.013***	-0.005**
	Females	209237	-332.3	-0.003***	-0.003***
Civil servant	Males	117980	-6854.5***	-0.019***	-0.005*
	Females	87047	-2696.1***	-0.012***	-0.003*
Production, low skill	Males	438953	-2360.5***	-0.015***	-0.004***
	Females	115456	-2332.2***	-0.016***	-0.007***

[18]Workers with small children have a quota of ten sick-leave days a year to take care of sick children. These absence spells are not covered by this dataset.

Trade & Commerce, low skill	Males	343102	-1944.2***	-0.014***	-0.005***
	Females	191815	-1525.9***	-0.013***	-0.007***
Age < 40, females	W/o children	248003	-1166.73***	-0.006***	-0.004***
	With children	463101	-718.8***	-0.006***	-0.003***
Age < 30	Males	165599	-2026.2*	-0.006	-0.002
	Females	111140	-286.7	-0.001	0.000
Age 30-40	Males	945512	-2634.1***	-0.013***	-0.004***
	Females	599964	-890.3***	-0.007***	-0.003***
Age 40-50	Males	835401	-3303.1***	-0.014***	-0.005***
	Females	557834	-1595.3***	-0.010***	-0.005***
Age 50-60	Males	714548	-3489.5***	-0.017***	-0.006***
	Females	458224	-1531.0***	-0.010***	-0.006***

Notes: Table 7 displays the results from estimation of (1.2) for various subgroups of workers. All results stem from the baseline specification of the model, reported in Column 3 in Tables 2 and 3. (The first stage equation (1.1) is also estimated separately for each sub-group but the results are not reported). The numbers of observations are reported for the linear model and are slightly lower for the log specification because workers with zero earnings are excluded. Robust standard errors clustered on physician are reported in brackets. Individual characteristics are age (linear, squared and cubed), years of education (linear, squared, cubed), low skill and high skill (dummy variables), marital status (3 dummies), gender, country of origin (2 dummies), sector of employment (11 dummies), calendar year (3 dummies).

The results from the linear model indicate that earnings of females with children are less affected than earnings of those without children. This difference disappears, however, when we estimate on log earnings. The estimates on employment are also fairly similar, slightly higher for those without children. This indicates that the reason why the linear estimate is lower for females with children, is partly that their employment is less affected and partly because they earn slightly less. These results suggest that we can reject the idea that the reason why females’ earnings are less affected by sick leave is related to children.

Ichino and Moretti (2009) explain that females’ earnings are less sensitive to sick leave than males’ earnings through statistical discrimination. Employers have a preconception of females having higher absence rates than males. [19] The result of this preconception is that they pay females less and make their earnings less dependent on sick leave. With time they get to know their employees better, and as they learn more about the individual they put less weight on their group-based prior, and as a consequence they treat men and women more similarly. A familiar argument, in the context of education, is made by Altonji and Pierret (2001) who hypothesize and find support for a claim that statistical discrimination decreases over time, as employers get to know their employees. The results for males and females in different age groups support this claim. The gender difference in how sensitive earnings are to sick leave is by far greatest for young workers. The estimated cost of a one percentage point increase in the sick-leave rate for males below 30 years is 0.6 percent, whereas the corresponding figure for females is 0.1 percent. None of these are statistically different from zero, however. The relative gap in the cost of absence decreases with age, for all but the oldest group. Also note that these gender differences are not driven by differences in the effects of sick leave on employment. This observation provides some support for the statistical discrimination hypothesis argued by Ichino and Moretti (2009).

5. Conclusion

In spite of 100 percent wage replacement during sickness, being on sick leave does not come without a cost. Using the leniency of primary care physicians, mandated to certify sick

[19] In Ichino and Moretti (2009) employers expect females to have higher sick-leave rates than males because of biological differences – the menstrual cycle.

leave, as an instrumental variable for individual sick leave, I find strong negative effects on subsequent earnings and employment. A one percentage increase in sick leave leads to a 1.2 percent reduction in earnings two years later. The probability of being employed is reduced by 0.5 percentage points. Consequently, the effects on employment can partly explain the effects on earnings, but not fully. Even for employees who remain employed there are negative effects on earnings, indicating that wages are affected. There are large differences between different groups of workers along dimensions such as: gender, job type, education and earnings. These differences are interesting in their own right because they shed light on wage setting mechanisms, implicit contractual arrangements and statistical discrimination in the labor market.

From a policy perspective the findings in this paper are worrying. The effects of sick leave on future earnings are probably difficult to avoid as they are subtly determined in a market. However, the negative effects of exogenous variations in sick leave on future employment are alarming. Sick leave is supposed to help workers recover in order to return to their jobs. These results indicate that being on sick leave can in some cases be a trap – from where the chance of escaping decreases by the day. Just from studying raw data or non-experimental settings it is well known that the fraction of workers returning to employment strongly decreases as the time spent on sick leave increases. However, this study is the first, as far as I know, to study this in a setting where causal effects can be identified. By comparing otherwise equal workers who differ only in the strictness of their primary physician, the causal effect of *being on sick leave* can be isolated from the underlying reasons that caused the sick leave. Physicians mandated to certify sick leave should think twice as the short-term solution to current health problems has serious and undesired consequences in the long term.

References

Angrist, J. D., Krueger, A. B., 2001: Instrumental variables and the search for identification: From supply and demand to natural experiments, *Journal of Economic Perspectives*, Vol.15, No 4, 69-85.

Angrist, J. D., Pischke, J., S., 2009: *Mostly harmless econometrics*, Princeton University Press, New Jersey.

Altonji, J. G. and Pierret, C. R., 2001: Employer Learning and Statistical Discrimination, *The Quarterly Journal of Economics,* February, 313-350.

Biørn, E., Godager, G., 2009: Does quality influence choice of general practitioner? An analysis of matched doctor-patient panel data, *Economic Modelling*, in press, available online.

Coles, M. G. and Treble, J. G., 1993: The price of worker reliability, *Economics Letters,* 41 149-155.

Coles, M. G. and Treble, J. G., 1996: Calculating the price of worker reliability, *Labour Economics,* 3, 169-188.

Dale-Olsen, H., 2006: Estimating Workers' Marginal Willingness to Pay for Safety Using Linked Employer-Employee Data, *Economica*, Vol. 73, 99-128.

Englund, L., Tibblin, G., Svärdskudd, K., 2000: Variations in sick-listing practice among male and female physicians of different specialities based on case vignettes, *Scandinavian Journal of Primary Health Care*, 18, 48-52.

Gorter, K., de Poel, S., de Melker, R., Kuyvenhoven, M., 2001: Variation in diagnosis and management of common foot problems by GPs, *Family Practice,* Vol. 18, No 6, 569-573.

Hansen, J., 2000: The effect of work absence on wages and wage gaps in Sweden, *Journal of Population Economics,* 13, 45-55.

Henningsen, M. and Hægeland, T., 2008: Downsizing as a sorting device: Are low-productive workers more likely to leave downsizing firms? Statistics Norway Discussion Papers, No. 543.

Henrekson, M., Persson, M., 2004: The Effects on Sick Leave of Changes in the Sickness Insurance System, *Journal of Labor economics*, Vol 22, 87-114.

Hesselius, P., 2007: Does sickness absence increase the risk of unemployment? *The Journal of Socio-Economics,* 36, 288-310.

Ichino, A., Riphan, R.T., 2005: The Effect of Employment Protection on Worker Effort: Absenteeism During and After Probation, *Journal of the European Economic Association*, Vol. 3, 120-143.

Ichino, A. and Moretti, E., 2009: Biological Gender Differences, Absenteeism and the Earnings Gap, *American Economic Journal: Applied Economics,* 1:1, 183-218.

Johansson, P., Palme, M., 2005: Moral hazard and sickness insurance, *Journal of Public Economics*, 2005, 1879-1890.

Manski, C. F., 1993: Identification of Endogenous Social Effects: The Reflection Problem, *Review of Economic Studies*, 60, 531-542.

Markussen, S. (2009), How physicians can reduce sick leave – Evidence from a natural experiment, working paper.

Markussen, S., Røed, K., Røgeberg and O.J., Gaure, S., 2009: The Anatomy of Absenteeism, IZA Discussion Paper No. 4240.

Nicholson, S., Pauly, M. V., Polsky, D., Sharda, C., Szrek, H. and Berger, M. L., 2006: Measuring the effects of work loss on productivity with team production, *Health Economics,* 15, 111-123.

Rosen, S., 1974: Hedonic Prices and Implicit Markets: Product Differentiation in Pure

Competition, *Journal of Political Economy* 82, 34–55.

Taylor, H. M., Karlin, S., 1998: *An Introduction to Stochastic Modelling*, 3rd edition, Academic Press Limited, London.

Weiss, A., 1985: Absenteeism and wages, *Economics Letters,* 19, 277-279.

Wooldridge, J., 2002: *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Appendix

Estimation of a physicians' leniency indicator in Markussen et al. (2009)

Using an extraordinarily rich set of Norwegian administrative data for 2001-2005, Markussen et al. (2009) estimates a leniency indicator for Norwegian primary care physicians. In their paper, individual sickness absence propensity is modeled by means of a multivariate hazard rate model. Since this paper makes direct use of some of the estimates obtained in their work, a description of their estimation strategy is provided, but for further details the reader is advised to read the full paper. In their model there are three possible states, $k=1,2,3,$ for each individual; attendance ($k=1$), absence with a minor disease ($k=2$) and absence with a major disease ($k=3$).[20] A present individual is can become absent due to either a minor or a major disease, and these events are modeled by means of a competing risks hazard rate model. Let K_1 be the set of feasible destination states for individuals currently in state 1, and let T_1 be the stochastic duration until one of the two possible events occurs. The competing hazards are then defined and specified as follows:

$$\theta_{1kit}\left(x_{it}\right) \equiv \lim_{\Delta t_1 \to 0} \frac{P\left(t_1 \leq T_1 \leq t_1 + \Delta t_1, K = k \mid T_1 \geq t_1, i\right)}{\Delta t_1} = \exp\left(x_{it}\beta_{1k}\right), k = 2,3$$

Where x_{it} is a vector comprising all observed explanatory variables assumed to affect individual *i'*s hazard rates at time *t*. Once absent, individuals are subject to the risk of recovery and hence of becoming present. Let $\{T_2, T_3\}$ be the stochastic durations of absence in states 2 and 3. The two single risk hazard rates are then defined and specified as follows:

[20] "Minor" and "major" diseases are grouped after the mean length of all spells for each diagnosis. If spells for a certain diagnosis on average last more than 16 days they are grouped as major, otherwise they are called minor.

$$\theta_{1kit}\left(x_{it},d_{it}\right) \equiv \lim_{\Delta t_j \to 0} \frac{P\left(t_j \le T_j \le t_j + \Delta t_j \mid T_j \ge t_j, i\right)}{\Delta t_j} = \exp\left(x_{it}\beta_{j1} + d_{it}\lambda_{j1}\right), j = 2,3$$

Where d_{it} is a vector describing the duration of an ongoing absence spell. The vector of explanatory variables x_{it} contains a wide range of potential absence determinants, such as age, gender, nationality, family situation, family background, important family events (pregnancy, divorce, death in the close family), place of residence, educational attainment, work-hours, earnings, tenure, local labor market conditions and calendar time. But, most importantly, x_{it} contains workplace and physician dummies for all physicians and workplaces with at least 100 employees. The richness of the data and the large number of observations are exploited to avoid unjustified functional form restrictions. This implies that virtually all the variables are dummy coded. For example, age is coded as a vector of 31 (time-varying) indicator variables (age=30,31,…,60), rather than as a polynomial in a single age-variable. Education is coded in the form of 65 dummy variables, reflecting both the length and type of education. Spell duration is coded by means of 28 dummy variables, allowing the piece-vice constant baseline hazards to differ before and after the 2004 reform of the Norwegian sickness insurance legislation (see Markussen, 2009 for a separate evaluation of this reform). For each of these periods, there are separate dummy variables for weeks 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11-26, 27-38, 39-49, 50-52. Calendar time is coded by means of quarterly dummy variables. A more detailed overview of explanatory variables is provided in the Appendix Table A1 in Markussen et al. (2009).

To derive the likelihood function for the observed data, each individual's event history is split into parts characterized by constant x_{it} and unchanged state (i.e., any changes in a explanatory variable or outcome triggers a new spell-part). Let S_{ji} (*j= 1,2,3*) be the set of observed spell parts in state *j* for individual *i*. Let l_{jis} denote the observed length of each of the spell parts $s \in S_{ji}$, and let the indicator variables $\left(y_{12is}, y_{13is}\right)$ denote whether a *state 1* spell part ended in a transition to *state 2* $\left(y_{12is}=1\right)$ or to *state 3* $\left(y_{13is}=1\right)$ or was censored $\left(y_{12is}=y_{13is}=0\right)$. Similarly, let $\left(y_{21is}, y_{31is}\right)$ indicate whether *state 2* and *state 3* spell-parts ended in work resumption or were censored. The likelihood function for individual *i* can then be written:

$$L_i = \prod_{s\in S_{1i}} \prod_{k\in\{2,3\}} \exp\left(-l_{1is}\left(\sum_{k\in\{2,3\}} \exp\left(x_{it}\beta_{1k}\right)\right)\right)\left[\exp\left(x_{it}\beta_{1k}\right)\right]^{y_{1kis}}$$
$$\times \prod_{j\in\{2,3\}} \prod_{s\in S_{ji}} \exp\left(-l_{jis}\left(\exp\left(x_{it}\beta_{j1} + d_{it}\lambda_{j1}\right)\right)\right)\left[\exp\left(x_{it}\beta_{j1} + d_{it}\lambda_{j1}\right)\right]^{y_{j1is}}$$

The number of spell parts in their analysis is approximately 50.5 million, and the number of coefficients as high as 35 000. For computational reasons they do not obtain standard errors for these parameters.

After estimation, the estimated coefficients are used to predict each employee's long-run absence propensity, defined formally as the absence rate that prevails as the time window goes towards infinity. These steady state absence rates are found by means of the limiting distribution of the Markovian transition matrix (Taylor and Karlin, 1998, p. 207) that can be constructed on the basis of each employee's four predicted hazard rates, taking into account the fact that recovery probabilities vary with absence duration. In this exercise, all explanatory variables (except spell duration) are held constant at the level prevailing at a particular point in time, implying that the steady state absence rates can be interpreted as the expected fraction of time of sickness absence over an infinitely long time horizon, given that no changes occur in individual characteristics or environmental factors. Let x_p denote the set of physician fixed effects and x_{-p} denote all the remaining observed characteristics in x. The four predicted hazards used for constructing the steady state absence rates can then be written as $\hat{\theta}_{ki}\left(x_{pi}, x_{-pi}\right) = \exp\left(x_{pi}\hat{\beta}_{jkp}\right)\exp\left(x_{-pi}\hat{\beta}_{jk-p}\right)$. To "remove" all sources of variation other than those caused by x_{pi}, the proportionality factors $\exp\left(x_{-pi}\hat{\beta}_{jk-p}\right)$ are replaced by their respective means, i.e. $\frac{1}{N}\sum_{i=1}^{N}\exp\left(x_{-pi}\hat{\beta}_{jk-p}\right)$. This implies that the steady state absence rates are a function of x_{pi} for workers who are representative along other dimensions. By construction, these rates are constant *within* physicians, such that we have a variable for each physician included in the estimation, which is the expected sick-leave rate for a worker conditional on nothing but his primary care physician. This variable is the strictness indicator used as an instrumental variable in Stage 1, denoted z_{it}.

Figure A1: Quality and leniency

Figure A1 compares the instrumental variable *physicians' leniency (z)* to age and gender adjusted mortality rates among the physicians' patients which are used as a proxy for *physicians' quality*. Physicians' quality is estimated on a dataset containing all Norwegian citizens aged 40 or above, for which the instrumental variable is available. On this dataset I fit a linear probability model. The endogenous variable is an indicator taking 1 if the person dies in the current year and zero otherwise. I explain individual mortality with gender (1 dummy), age (69 dummy variables), year (4 dummy variables) and physician (3205 physicians). What are referred to as physicians' quality in Figure A1 are the coefficients for physicians in this regression. The slope of the regression line is not significantly different from zero and the coefficient of correlation is as low as 0.0125.