DISCUSSION PAPER SERIES

IZA

IZA DP No. 4240

The Anatomy of Absenteeism

Simen Markussen
Knut Røed
Ole J. Røgeberg
Simen Gaure

June 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Anatomy of Absenteeism

Simen Markussen
Ragnar Frisch Centre for Economic Research

Knut Røed
Ragnar Frisch Centre for Economic Research
and IZA

Ole J. Røgeberg
Ragnar Frisch Centre for Economic Research

Simen Gaure
Ragnar Frisch Centre for Economic Research

Discussion Paper No. 4240
June 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

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

IZA Discussion Paper No. 4240
June 2009

ABSTRACT

The Anatomy of Absenteeism*

Based on comprehensive administrative register data from Norway, we examine the determinants of sickness absence behavior; in terms of employee characteristics workplace characteristics, panel doctor characteristics, and economic conditions. The analysis is based on a novel concept of a worker's steady state sickness absence propensity, computed from a multivariate hazard rate model designed to predict the incidence and the duration of sickness absence for all workers. Key conclusions are i) that most of the cross-sectional variation in absenteeism is caused by genuine employee heterogeneity; ii) that the identity of a person's panel doctor has a significant impact on absence propensity; iii) that sickness absence insurance is frequently certified for reasons other than sickness; and iv) that the recovery rate rises enormously just prior to the exhaustion of sickness insurance benefits.

JEL Classification: C14, C41, H55, I18, J22

Keywords: sickness absence, multivariate hazards, MMPH, NPMLE

Corresponding author:

Knut Røed
The Ragnar Frisch Centre for Economic Research
Gaustadalléen 21
0349 Oslo
Norway
E-mail: knut.roed@frisch.uio.no

* This paper is part of the project "A viable welfare state" financed by the Norwegian Research Council. One of the authors (Ole Røgeberg) has also received financial support from the Health Economics Research Programme at the University of Oslo - HERO. We wish to thank Erik Biørn, Harald Goldstein, and participants at the ESPE 2009 congress in Seville for comments.

1. Introduction

It is a well known fact that the level of sickness absence varies substantially across time and space; see Bonato and Lusinyan (2004). Economists have tended to focus on the financial incentives provided by sickness insurance institutions and their interaction with employment protection legislation and cyclical fluctuations when explaining trends and disparities in absenteeism; see, e.g., Henrekson and Persson (2004), Ichino and Riphahn (2005), Ruhm (2000), and Johansson and Palme (2002) for recent evidence. Yet, although the disincentive effects arising from social insurance have been convincingly established empirically, policy makers in many welfare state economies remain hesitant towards fundamental reform. Apparently, the welfare gains associated with income security and equality are considered sufficient to justify the costs arising from a higher level of absenteeism.

This statement certainly applies for Norway. On a typical working day, around seven per cent of Norwegian employees are absent from work due to sickness. Their insurance coverage is 100 percent of regular earnings from the first day of absence. In total, the resultant insurance payments amount to approximately 2.4 per cent of the Norwegian GDP. Hence, the costs associated with sickness absence – in terms of forgone labor supply as well as direct insurance payments – are substantial. Yet, a reduction of the replacement ratio is not on the political agenda. This does not imply that the problems associated with high absenteeism pass unrecognized. But rather than trading off lower absenteeism against poorer insurance coverage, policy makers have aimed at shifting the tradeoff locus itself. Their primary means have been to overhaul the absence certification regulations and to encourage firms to make workplace environments more "inclusive". These policies have been motivated by the observations that more than 85 percent of sickness absence is certified by a physician, and that absenteeism varies a lot across time, workplaces, and panel doctor patient lists, despite the lack of variation in financial incentives. Figure 1 illustrates the point. The upper panel shows the evolvement of aggregate certified (lasting more than three days) and non-certified absence rates over time from 2001 through 2005. Absenteeism apparently trended slightly upwards until 2004, after which the certified absence rate declined sharply. The conspicuous decline in 2004 coincided with a revision of the guidelines regulating the physicians' certification of absence spells, emphasizing that sickness is not a sufficient conditions for absence and that *activity* is normally preferable to *rest* during longer-term sickness periods.

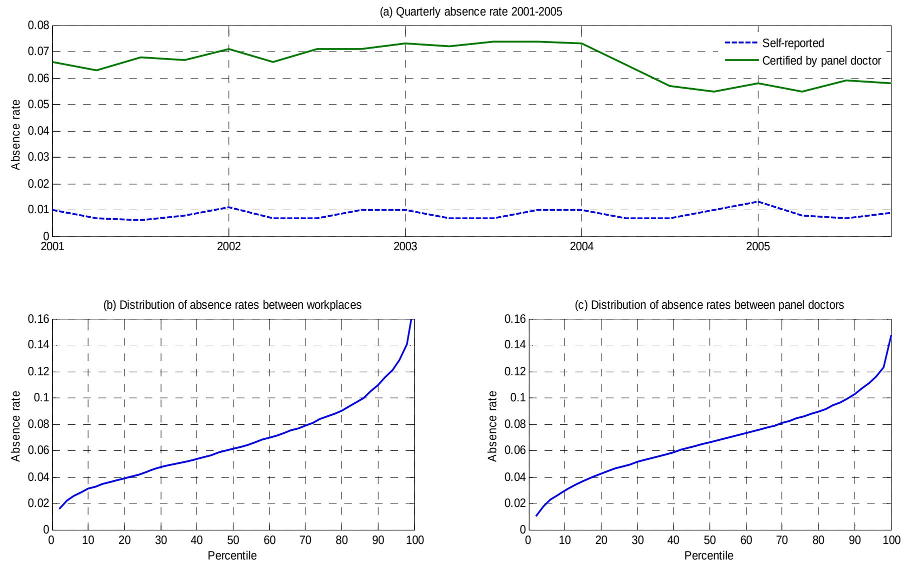

Figure 1: Absenteeism in Norway: Panel (a) displays aggregate quarterly absence rate 2001-2005. Panel (b) displays the distribution of mean certified absence rates 2001-2005 within firms (>100 employees) and panel doctor patient lists (>100 employed patients).
Sources: Panel a: Statistics Norway; Panels b and c: Own calculations based on merged register data (see Section 2 of this paper).

The lower panels show the distributions of mean certified absence rates over all these five years by workplaces and panel doctors (conditional on at least 100 employees or 100 employed patients), respectively. The variation is indeed substantial. For example, while workplaces in the 10th percentile of the firms' absence rate distribution on average had absence rates around 3.1 percent, workplaces in the 90th percentile had absence rates around 11.0 percent. The corresponding numbers for panel doctors were 3.0 and 10.3 percent. Now, absence variation across time and space reflects a combination of at least three types of factors: i) genuine randomness, ii) nonrandom sorting of employees and common confounding factors, and iii) causal impacts of the factors in question. In order to identify the potential for policy intervention it is necessary to disentangle the contributions from each of these components.

The aim of this paper is to examine the origins of the observed variation in absenteeism in Norway across time and space. Norway is an interesting case to study for a number of reasons; the absence rate is extremely high and variable, the sickness insurance system is extremely generous, and the available data are extremely informative. In this study we use complete register data for all certified absence spells in Norway – their starting dates, their stopping dates, and their outcomes – from 2001 through 2005. The starting point of our dataset corresponds to the introduction of the panel doctor system in Norway, whereby each citizen was assigned to a single General Practitioner (GP). Our dataset includes the linkage between citizens and GP's on a monthly basis, as well as information about citizens and physicians. It also includes the linkage between workers and workplaces. The data allow us to examine the sources of variation in sickness absence propensity due to individual factors (gender, age, family situation, work-hours, tenure, nationality, education/occupation, income, social background, wealth, family events, etc.), workplace-characteristics (industry, size, turnover, downsizing, employee-composition), the local economic environment (job-finding rate, employment rate), panel doctor characteristics (age, gender, specialization, number of patients, deviation from desired number of patients, degree of competition), time trends, seasonal fluctuations, and institutional characteristics (system reform).

Although there is a vast existing literature regarding various determinants of sickness absence behavior, we are not aware of any previous studies aiming at a comprehensive quantitative decomposition of the kind offered in this paper. This also implies that appropriate decomposition tools need to be developed. Hence, the paper provides novel contributions both to the methodology of sickness absence decomposition and to the more substantive issue of quantifying the key determinants of sickness absence behavior in a modern welfare state. We model individual absence behavior by means of a multivariate hazard rate model, accounting for the incidence as well as the recovery from two different types of absence spells – minor and major – distinguished on the basis of medical diagnosis. The model is estimated by means of the nonparametric maximum likelihood estimator (NPMLE). The resultant predicted hazard rate profiles are subsequently used to compute each employee's steady state absence rate, i.e., the absence rate that can be expected to prevail in the long run.

Our main findings are as follows: First, the longitudinal variation in absenteeism observed between 2001 and 2005 is to a limited extent explained by sorting into and out of the workforce. Individual absenteeism is changeable, and the 2004 reform in the absence certification regulations – with larger emphasis on work-attendance during sickness episodes – did cause a significant drop in absence rates. Second, even though roughly half of the variation in absenteeism between workplaces is accounted for by employee sorting, substantial workplace-differences remain. Workplace environments do have a considerable impact on absenteeism, and we find indications that differences between workplaces to some extent are amplified by social interaction processes among colleagues. Third, the variation in absenteeism between employees listed with different panel doctors does not primarily result from systematic employee sorting. The upshot is that panel doctors' certification practices vary substantially, and that these practices are important for actual absence decisions. Fourth, in accordance with existing evidence, we find that absence behavior is responsive towards employees' financial incentives. In particular, we identify a dramatic rise in recovery rates when the generous sickness benefit insurance is exhausted after one year of absence. Finally, although we show that observed employee characteristics – such as age, gender, social background, family situation, education, and occupation – all have substantial impacts on absence behavior, we conclude that unobserved differences in individual absence propensities are extremely important for understanding the observed cross-sectional variation in absence behavior. In the long run, we find that such unobserved factors explain as much as two thirds of the overall variance in individual absence rates.

The next section gives a brief description of the data and of the institutional circumstances from which they are generated. Section 3 presents our empirical methodology, Section 4 presents the estimation results, and Section 5 translates the estimation results into a distribution of steady state absence rates and provides a variance decomposition of these rates. Section 6 concludes.

2. Data and institutions

The data we use in this paper comprise starting dates and stopping dates for all certified sickness absence spells in Norway during the period from June 2001 to December 2005. They also include the medical reason for each absence spell (diagnosis), the (encrypted) identity of the physician responsible for its certification, and the (encrypted) identity of each citizen's panel doctor.[1] The data on absence spells are merged with a number of other administrative data registers providing detailed information about individual employees, their panel doctors, their workplaces, and the institutional and economic environments they face.

The starting point of our empirical analysis is the population at risk of becoming absent from work in June 2001 – one month after the panel doctor reform was implemented in Norway. To start with, our analysis population consists of all employed individuals in Norway aged 30-60 at this point in time. After that, new individuals are in-

[1] A large fraction of absence spells are (initially) certified by a physician other than the individual's own panel doctor (e.g., by casualty units). However, to the extent that the absence spell lasts more than a few days, certification renewals will typically be taken care of by the panel doctor.

cluded in the dataset as they become 30 years and/or become employed. Individuals are removed from the dataset (censored) as they become 61 years or non-employed. Hence, at any point in time during 2001-2005, our analysis population consists of all employed individuals in Norway aged 30-60. In total 1.78 million individuals – and 3.7 million absence spells – are included.

Norwegian employees are normally paid their regular salary during sickness absence for up to one year; i.e., there is a 100 percent replacement ratio. During the first 16 days of absence, the expenses are covered by the employer, after which the social security system foots the bill.[2] The general rule is that absence spells lasting more than three days must be certified by a physician. However, certification is not required until the 9th day for employees in firms participating in a tripartite "inclusive workplace agreement" (IWA) between employers, employees and the state.[3] Within the framework of this same agreement, new absence certification regulations were implemented (for everyone) in July 2004, implying that *partial* sickness absence became the "default" option after 8 weeks of sickness absence. Sickness benefits cannot be paid for more than a year. After that, the claimant is referred to the much less generous rehabilitation benefit (with replacement ratio around 66 percent).

3. Empirical strategy

This section describes our methodology. We start out by presenting the statistical model that we use to account for sickness absence behavior, conditional on a vector of time varying observed covariates x_{it} and a vector of time-invariant unobserved covariates v_i. We then explain how we estimate the parameters of that model by means of a nonparametric maximum likelihood estimator (NPMLE). Finally, we describe how we use the estimated model for simulation purposes.

1.1 The model

We model individual sickness absence propensity by means of a multivariate hazard rate model. The model distinguishes between "minor" (acute) and "major" (potentially chronic) diseases, based on the recorded diagnosis. The distinction is made on the basis of the aggregate duration distribution of absence spells *by diagnosis*, such that diagnoses

[2] There is an upper ceiling on the sickness insurance benefits paid out be the social security system (corresponding to a yearly income of around 50,000 USD), but employers typically cover the wedge between the maximum social security payment and normal earnings..

[3] Frequently absent employees need certification from the first day of absence.

with mean durations below 17 days are classified as minor, while diagnoses with longer mean durations are classified as major. According to this classification, the group of minor diseases is dominated by respiratory infections, virus diseases, and gastrointestinal diseases, while the group of major diseases is dominated by musculoskeletal and mental diseases. The resultant distributions of absence spells are described in Table 1 and in Figure 2. Around 72 percent of the absence spells are classified as major, and given their longer expected durations they account for as much as 93 percent of overall certified absence. While 80 percent of the minor spells lasts shorter than two weeks, this is the case for around 55 percent of the major spells. Roughly 10 percent of the major spells last longer than five months. Our classification does not correspond perfectly to actual disease "seriousness", however, and some minor spells turn out to be long-lasting, while some major spells turn out to be short.

Table 1. Sickness absence spells- minor and major diseases

	Minor	Major
Number of spells 2001-2005	1 052 611	2 666 316
Mean length (days)	8.9	46.1
Share of spells (%)	28.3	71.7
Share of total absence days (%)	7.1	92.9
Share right-censored spells (%)	0.44	2.05

Note: Spells are right-censored if they are terminated without work-resumption. This typically happens after 12 months of absence, at death/emigration, and at the end of the observation period.

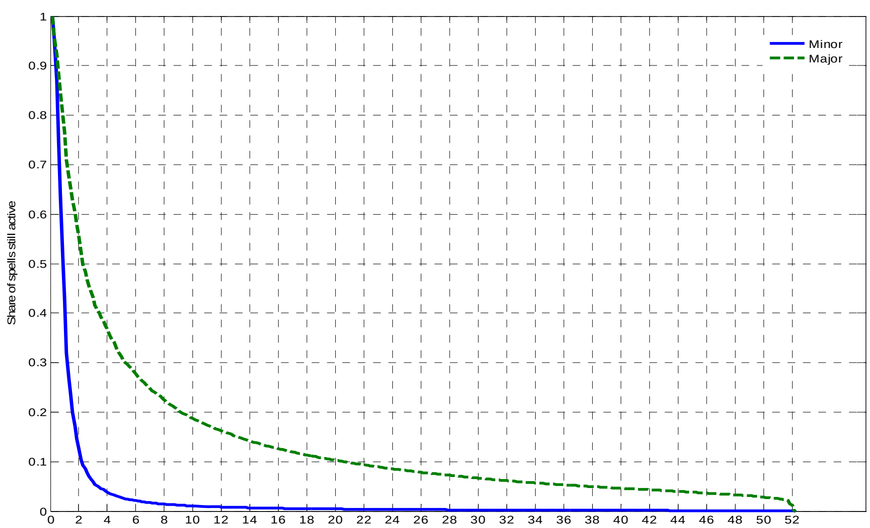

Figure 2: Survival curves for minor and major absence spells (life table estimators).

There are three alternative states, $k=1,2,3$, that an individual can occupy in our model; attendance (k=1), absence with a minor disease ($k=2$) and absence with a major disease (k=3). A present individual is under risk of becoming absent due to either a minor or a major disease; hence we model these events 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 occur. The competing hazards are then defined and specified as follows:

$$\theta_{1kit}(x_{it}, v_{1ki}) \equiv \lim_{\Delta t_1 \to 0} \frac{P(t_1 \le T_1 \le t_1 + \Delta t_1, K = k \mid T_1 \ge t_1, i)}{\Delta t_1} = \exp(x_{it}\beta_{1k} + v_{1ki}), \ \ k = 2,3, (1)$$

where x_{it} is a vector comprising all observed explanatory variables assumed to affect individual *i's* hazard rates at time *t* and (v_{12i}, v_{13i}) are time-invariant unobserved employee characteristics.

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, respectively. The two single risk hazard rates are then defined and specified as follows:

$$\theta_{j1it}(x_{it}, d_{it}, v_{j1i}) \equiv \lim_{\Delta t_j \to 0} \frac{P(t_j \le T_j \le t_j + \Delta t_j \mid T_j \ge t_j, i)}{\Delta t_j} = \exp(x_{it}\beta_{j1} + d_{it}\lambda_{j1} + v_{j1i}), \ j = 2,3, (2)$$

where d_{it} is a vector describing the duration of an ongoing absence spell and (v_{21i}, v_{31i}) are time-invariant unobserved employee characteristics.

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 close family), place of residence, educational attainment, workplace characteristics, industry, work-hours, earnings, tenure, local labor market conditions, panel doctor characteristics, and calendar time. We exploit the richness of our data and the large number of observations 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 the 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. In 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. And calendar time is coded by means of quarterly dummy variables. A more detailed overview of explanatory variables is provided in Appendix Table A1.

1.2 Estimation

To derive the likelihood function for observed data, we split each individual's event history into parts characterized by constant x_{it} and unchanged state (i.e., any change in any 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 part $s \in S_{ji}$, and let the indicator variables (y_{12is}, y_{13is}) denote whether a state 1 spell part ended in a transition to state 2 $(y_{12is}=1)$ or to state 3 $(y_{13is}=1)$ or was censored $(y_{12is}=y_{13is}=0)$. Similarly, let (y_{21is}, y_{31is}) indicate whether state 2 and state 3 spell-parts ended in work resumption or were censored. Conditional on unobserved heterogeneity, the likelihood function for individual *i* can then be written

$$\begin{aligned} L_i(v_i) = & \prod_{s\in S_{1i}} \prod_{k\in\{2,3\}} \exp\left(-l_{1is}\left(\sum_{k\in\{2,3\}} \exp(x_{it}\beta_{1k}+v_{2ki})\right)\right)\left[\exp(x_{it}\beta_{1k}+v_{1ki})\right]^{y_{1kis}} \\ & \times \prod_{j\in\{2,3\}} \prod_{s\in S_{ji}} \exp\left(-l_{jis}\left(\exp(x_{it}\beta_{j1}+d_{it}\lambda_{j1}+v_{j1i})\right)\right)\left[\exp(x_{it}\beta_{j1}+d_{it}\lambda_{j1}+v_{j1i})\right]^{y_{j1is}} \end{aligned} \tag{3}$$

where $v_i=(v_{12i}, v_{13i}, v_{21i}, v_{31i})$. The total number of spell parts included in our analysis is around 50.5 million.[4]

Since the likelihood contribution in (3) contains unobserved variables, it cannot be used directly for estimation purposes. This problem may be solved by formulating a model for the joint distribution of unobserved heterogeneity and replace Equation (3) with its expectation. In order to avoid unjustified assumptions, we approximate unobserved heterogeneity by means of a discrete distribution. Let *Q* be the (the a priori unknown) number of support points in this distribution and let $\{v_l, p_l\},\ l=1,2,...Q,$ be the associated location vectors and probabilities. In terms of observed variables, the likelihood function is then given as

$$L=\prod_{i=1}^{N} \underset{v_i}{E}[L_i(v_i)] = \prod_{i=1}^{N}\sum_{l=1}^{Q} p_l L_i(v_l), \quad \sum_{l=1}^{Q} p_l = 1, \tag{4}$$

where $L_i(v_l)$ is given in Equation (3). Our estimation procedure is to maximize the likelihood function (4) with respect to all the model and heterogeneity parameters repeatedly

---

[4] At the start of our data window in 2001 some employees are already absent due to sickness. These spells are left out of the analysis, and the individuals in question are included when/if they again become present. The reason for this is that exploitation of ongoing spells would involve some rather intricate initial conditions problems, since the initial condition in this case not only comprises a particular state, but also a particular duration.

for alternative values of *Q*. The non-parametric maximum likelihood estimators (NPMLE) are obtained by starting out with *Q=1*, and then expanding the model with new support points until the model is "saturated" in the sense that it is no longer possible to increase the likelihood function by adding more points (Lindsay 1983; Heckman and Singer 1984). At each stage of the estimation process, we examine the appropriateness of an additional mass-point by means of simulated annealing (Goffe, Ferrier, and Rogers 1994). The preferred model is then selected on the basis of the Akaike Information Criterion (AIC). Monte Carlo evidence presented in Gaure, Røed and Zhang (2007) indicates that parameter estimates obtained this way are consistent and approximately normally distributed. They also indicate that the standard errors conditional on the optimal number of support points are valid for the unconditional model as well, and hence can be used for standard inference purposes.

An implicit assumption in this model is that movements into and out of employment (and, hence, into and out of the analysis population) are exogenous with respect to the two modeled hazard rates, conditional on all observed explanatory variables. This assumption is probably violated. However, the extraordinary rich set of observed characteristics used in this analysis should ensure that the potential biases arising from this violation are reduced to a minimum.

1.3 Simulation

Once the model is estimated, it can be used for simulation purposes, both as a means to assess the model's performance and as a tool for investigating the influence of particular variable groups on aggregate absenteeism. We generate *simulated* absence data by equipping each employee with his/her actual observed explanatory variables at the moment of entry into the dataset, make a drawing of the unobserved variable vector from the estimated joint heterogeneity distribution, and then let the future movements across states be determined by a sequence of lotteries. The probabilities entering into these lotteries are computed from the four predicted hazard rate profiles. These profiles are partly determined by the worker's characteristics and partly by the economic environment, e.g., through the evolvement of calendar time dummy variables. In order to predict each worker's long-term (steady state) absence rate, we also perform simulations based on a completely static infinite horizon environment. These simulations are more thoroughly explained in Section 5.

4. Estimation results

This section presents the estimation results. All the results are based on the preferred mixture model, for which the unobserved heterogeneity distribution ended up having 29 distinct support points.[5] In total, the chosen model contains 1,702 parameters to estimate; hence we cannot present the results in any detail. We limit ourselves to presenting some key findings believed to be of general interest. Given our extensive usage of indicator variables, standard errors are typically of limited interest, since they essentially measure the statistical uncertainty relative to an arbitrarily selected reference group. Hence, for expository reasons we do not report standard errors or confidence intervals except were this is deemed to be of particular interest. The size of our dataset ensures that most parameters are estimated with great precision, and we will argue that the focus should really be on *substantive*, rather than *statistical* significance. Note also that many of the results are presented directly in the text, without reference to a table. Complete estimation results – with standard errors – are accessible from our web page www.frisch.uio.no/docs/absenteeism.html.

1.4 The longitudinal variation in absenteeism - time and duration dependence

We start out by assessing the estimated model's ability to reproduce the longitudinal variation in actual certified absence rates and to examine the extent to which the changes in aggregate absenteeism over time can be explained by sorting into and out of the workforce. The upper panel of Figure 3 compares actual monthly absence rates over time to the aggregate monthly absence rates obtained by simulating the employees' absence behavior on the basis of the estimated model. We consider the model fit to be reasonably good, except that it fails to fully capture the seasonal pattern in absenteeism. The reason for this deficiency is that calendar time is represented in our model by quarterly, rather than monthly, dummy variables. Note that the low absence rates in the beginning of the data period reflect that we have conditioned on all employees being present at work to start with. This implies that aggregate absence rates are only comparable over time starting from June 2002.

[5] None of the results presented in this paper are sensitive with respect to the exact number of support-points in the heterogeneity distribution.

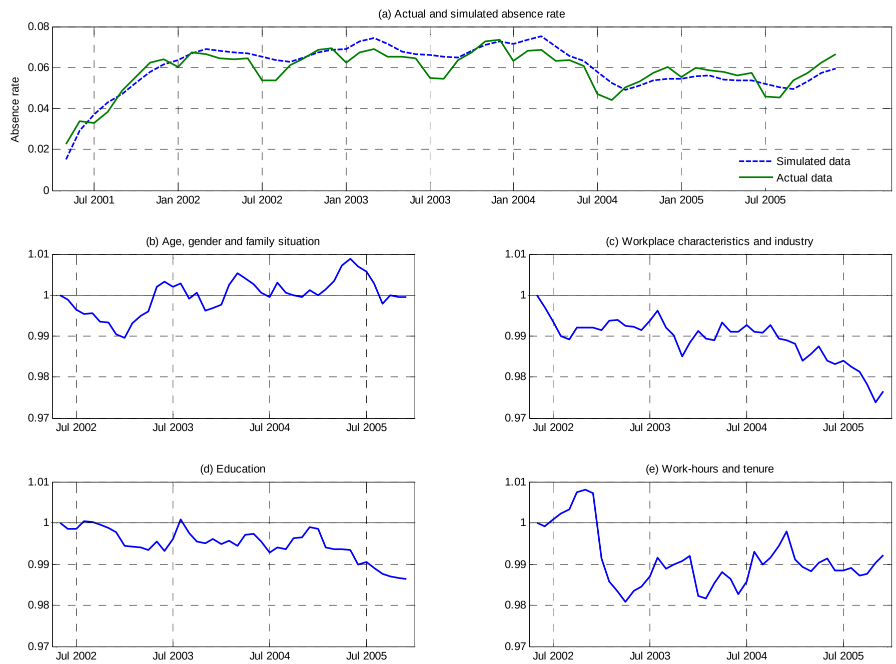

Figure 3. Actual and simulated absence rates 2001-2005 (panel (a)) and the estimated contribution to relative changes in the aggregate absence rate by four different variable groups (panels (b)-(e)).
Note: Panels (b)-(e) report relative changes caused by the variable groups in question. The curves are normalized to unity in June 2002.

Changes in aggregate absenteeism may result from changes in employees' absence behavior and/or from changes in the composition of workers and workplaces. The four lower panels in Figure 3 illustrate the isolated composition-effects arising from some selected variable groups.[6] They show that demographic changes contributed to a small (1 percent) decline in the aggregate absence rate during 2002, and to a small rise afterwards. Changes in work-hours also contributed to a decline in absenteeism during the economic downturn in 2002, while changes in the education and industry composition contributed to a somewhat larger and more trend-like decline. The key message coming out of this exercise, however, is that impacts of compositional changes are small over the time-horizon covered by our analysis. Thus, the changes that occurred in absenteeism over these five years primarily reflected variation in each workers' absence behavior, and not changes in the composition of workers and workplaces.[7]

[6] These profiles are obtained by simulating absence behaviour when only the variable group in question is allowed to vary over time, and then normalize the absence rate to unity in May 2002.

[7] Note that we also examine the cyclical sorting-hypothesis by including the local (municipality) employment rate as an explanatory variable in the model. A high local employment rate may be taken as an indicator that individuals with poor health have been included in the workforce; hence we may expect high employment to coexist with high absence rates. Our findings suggest that a higher local employment rate implies both higher entry and recovery rates, indicating that the marginal members of the workforce have

Figure 4 displays the estimated calendar time effects for each of the four hazard rates in our model (exp of coefficients attached to the quarterly time dummy variables). The entry rate into minor diagnosis is completely dominated by the seasonal pattern, and no obvious time-trend can be spotted. For the entry rate to major diagnosis, however, there appears to have been a downward shift from the third quarter of 2004, coinciding with the reform of the sickness absence certification regulations. Recovery rates (the two lower panels) also seem to have risen slightly after this reform.

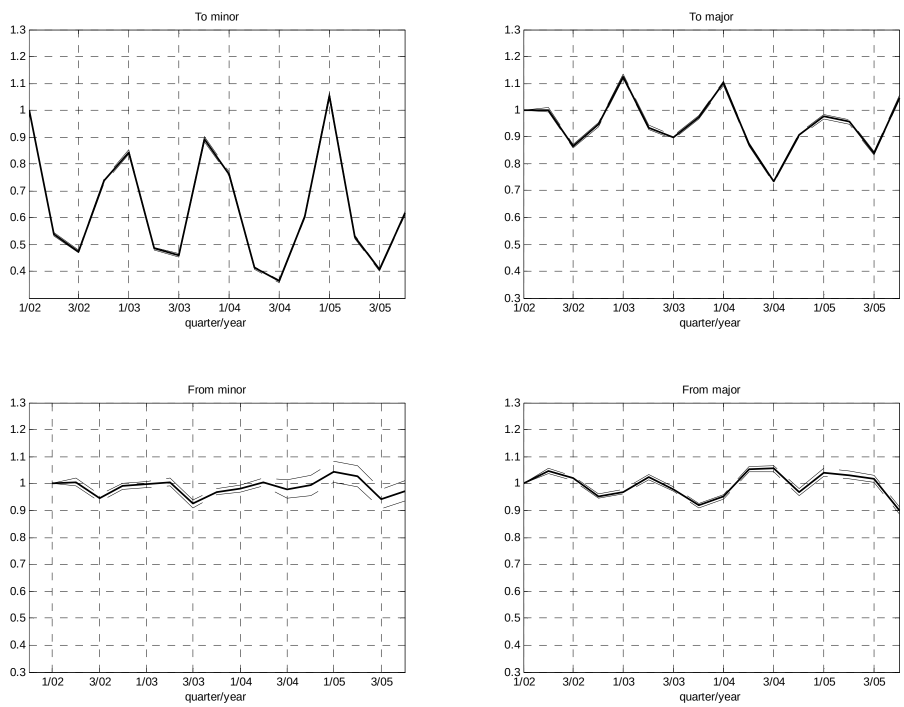

Figure 4. Estimated relative impacts of calendar on the four hazard rates (with 95 percent confidence intervals; reference is 1st quarter 2002).

The calendar time developments in the recovery rates shown in Figure 4 are evaluated at the beginning (first week) of absence spells. In order to assess the potential impacts of the 2004 reform on the *recovery profiles* (the duration dependence in the recovery hazards), we have estimated the two recovery baseline hazard rates separately before and after the reform. The result is illustrated in Figure 5, where we have scaled the estimated recovery profiles such that they start out at a level corresponding to the ob-

---

more frequent – but also less serious – spells of sickness absence. We use simulations to assess whether higher employment increases or decreases the average absence rate, i.e. whether the positive entry or the negative recovery rates dominate. We find that an increase in the local employment rate from 80 to 85 percent *reduces* the average absence rate by around 0.3 percentage points. Hence, our results do not confirm the idea that high employment entails a high rate of sickness absence.

served average recovery frequency during the first absence week. Recall that the reform made work attendance the “default” option after 8 weeks of sickness absence; hence we would expect it to raise recovery rates after 8 weeks. Whereas the recovery rates from the minor diseases are virtually unchanged, there is indeed a slight increase in the relative recovery probabilities at longer durations for the major disease diagnosis after the reform (although it is hardly visible in the graph, the shift is statistically significant at conventional levels). However, simulations show that the quantitative importance of this shift on the aggregate absence rate is modest, around 0.2 percentage points (3.3 percent). In total, the reform appears to have shifted the absence rate downwards with approximately one percentage point; see Markussen (2009) for a separate analysis of this issue.

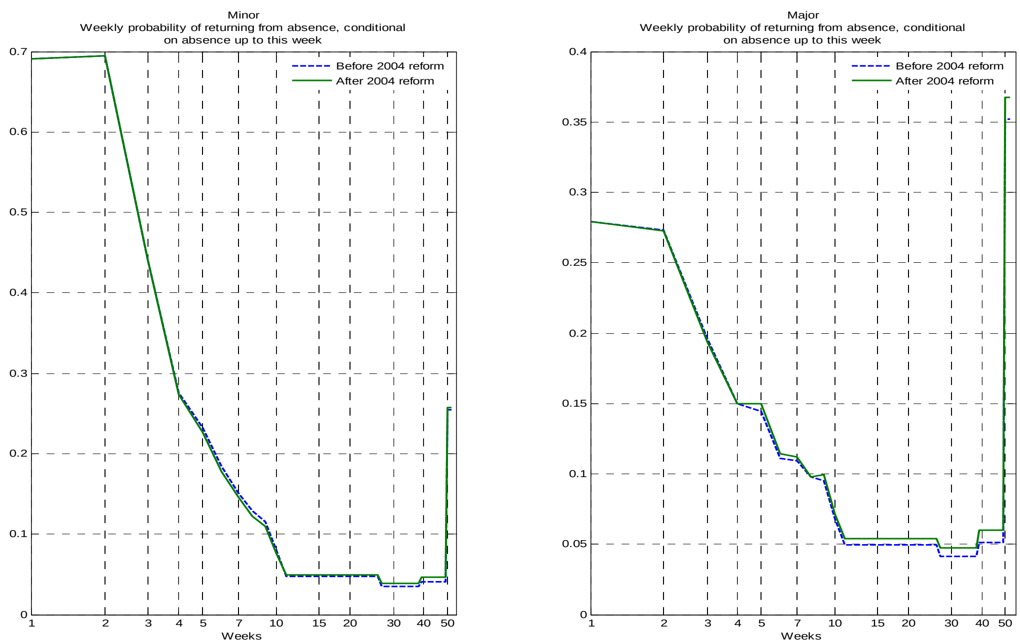

Figure 5. Duration dependence in recovery hazards. Before and after the reform in 2004 (log-scale on the horizontal axes).
Note that the scales on the vertical axes differ between the two graphs.

The baseline hazards depicted in Figure 5 illustrate that there is strong negative duration dependence in recovery prospects. For the average employee, the probability of recovering from a major disease declines from around 28 percent during the first week to around 5 percent after 10 weeks. As the sickness insurance period approaches exhaustion after one year, the recovery hazard again increases sharply, and the weekly probability of returning from a major disease absence spell is never larger than during the last four weeks. The spikes in the recovery hazard just prior to sickness benefit exhaustion is reminiscent of the benefit exhaustion spike frequently encountered in unemployment duration analyses; see, e.g., Card, Chetty, and Weber (2007) for a recent review of the literature, and Røed and Zhang (2003) for Norwegian evidence. Even though an extremely small fraction of absentees is directly affected by sickness benefit exhaustion (see Figure 2), this finding indicates that financial incentives have the potential of shifting absence behavior quite substantially.

1.5 Age, gender, and family background

We now take a closer look at how various groups of explanatory variables affect entry and recovery hazard rates. The partial impacts of age are illustrated in Figure 6. The probability of entering into a sickness absence spell declines sharply with age up to around 45 years. A 30 year employee has a 70 percent higher entry rate to minor disease absence and a 10 percent higher entry rate to major disease absence than a 45 year old employee, *other things equal*. Above 45 years, the entry rates either stabilize (minor diseases) or rise (major diseases). The probability of recovering from an illness declines monotonously with age. Taken together, these estimates imply that the overall minor disease absence propensity *declines* with age up to around age 50, while major disease absence rises monotonously and significantly with age. Given that individuals' health conditions are typically assumed to deteriorate with age, we find the extremely high absence *entry rates* among young workers intriguing.[8] To the extent that absenteeism represents a withdrawal from unsatisfactory work conditions, the declining absence frequency by age may reflect that older employees have had more time to find a satisfactory job match (Martocchio, 1989). Moreover, systematic sorting out of the labor force *by age* probably ensures that the most absent adolescents are no longer in the work force when they reach their 40's. However, our findings may also reflect that young workers are bearers of a new and less strict norm set, and hence have lower thresholds for claiming sick.

[8] According to the Norwegian level of living sample survey in 2005 (provided by Statistics Norway), 88 percent of individuals aged 25-44 characterize their own health as good or very good. The same applies for 76 percent of individuals aged 45-66.

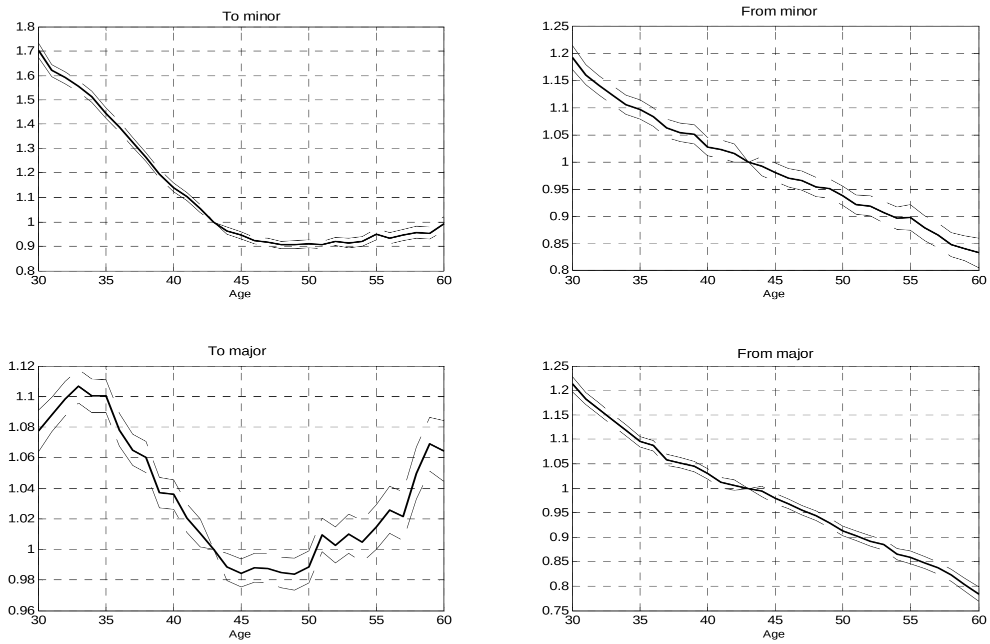

Figure 6. Estimated relative impact of age on the four hazard rates (with 95 percent confidence intervals; reference is age=43)

Table 2 presents the estimated impacts of gender interacted with family situation. Women have much higher entry rates than men, regardless of whether we compare married, separated/divorced or single employees, with or without children. Women's entry rates to certified sickness absence exceed those of men by as much as 45-68 percent for minor diseases and 26-43 percent for major diseases.[9] The gender differentials in recovery rates are much smaller. Women tend to recover somewhat slower than men from major disease absence, particularly if they have children.

Table 2. The impact of being female rather than male (percentage change in hazard rate)

	Absence entry		Recovery	
	To minor	To major	From minor	From major
Being a female, by marital status and parenthood (reference=male)				
Married				
Without children	53.2	26.1	3.6	3.2
With children	45.5	31.5	1.4	-3.7
Separated/divorced				
Without children	67.4	38.1	3.4	4.5
With children	63.9	43.3	2.3	0.5
Never married				
Without children	57.6	43.1	2.4	-4.7
With children	57.3	35.9	2.0	-2.4

[9] Note that these differences cannot be explained by either the menstrual cycle (Ichino and Moretti, 2009) or by pregnancies. Menstrual pain is recorded as a separate diagnosis, and this diagnosis accounts for less than 1 percent of the certified minor disease absence among females. Pregnancies are controlled for through separate variables; see Section 4.6.

There are large differences in absence patterns across different nationalities. Immigrants from EU or USA have entry rates to absence around 9 percent above those of natives. Immigrants from outside EU/USA have entry rates that are as much as 43 (minor) and 58 (major) percent higher than natives, ceteris paribus. Recovery rates are fairly similar across nationalities, with the exception that immigrants from outside EU/USA have 18 percent higher recovery rates from major diseases than natives. The latter finding reflects a general pattern that a high recovery rate is a mirror image of a high entry rate, probably indicating that a high entry rate signals low absence thresholds.

In accordance with existing evidence (see, e.g., Marmot, 2004, for a recent review), we find a strong social gradient in absenteeism. Even conditional on own education, job type, income, and wealth, family background has a significant impact on the probability of entering into a sickness absence spell. Workers borne in families with at least one parent in the highest education bracket (PhD) have around nine percent lower probability of starting a minor disease absence spell and almost 18 percent lower probability of starting a major disease absence spell than an otherwise identical person born in a family with both parents in the lowest education bracket (only compulsory education). Income differences in the parent generation reinforce this social gradient, as the offspring’s rate of entry into major disease spells declines significantly with the parents’ income. The parents’ health condition also seems to be transferred across generations. Indicators of early death or disability of parents predict high offspring absence rates. For example, with both parents being disabled, the offspring’s entry rate into a major disease spell rises by 25 percent, ceteris paribus.

1.6 Human capital and job characteristics

Table 3 presents selected estimates regarding the association between own educational attainment and absence behavior. Educational attainment sharply reduces the rates of entry into both minor and major disease absence. And it seems to be the *level* of education that matters – not its type. This pattern is also evident for a number of education tracks that are not reported in the table.

Table 3. The impact of education (percentage change in hazard rates)

	Absence entry		*Recovery*	
	To minor	*To major*	*From minor*	*From major*
Educational attainment; selected estimates (reference= compulsory education only)				
General educations				
Uncompleted high school	-15.2	-21.8	3.5	3.4
Completed high school	-25.3	-34.8	5.5	2.9
Health related educations				
Uncompleted high school	-3.5	-3.8	2.8	1.3
Completed high school	-2.6	-4.1	2.5	3.2
College, lower level (nurses)	-23.6	-25.4	0.8	-6.1
University, higher level (doctors)	-55.0	-59.6	5.7	-9.3
PhD	-68.5	-71.7	-0.8	-7.8
Technical/mechanical education				
Uncompleted high school	2.3	-0.2	1.6	1.7
Completed high school	-8.7	-15.9	2.3	3.6
High school w/ext.	-23.6	-35.8	3.5	0.7
College, lower level	-40.2	-57.8	7.9	3.2
College/University, higher level	-49.4	-69.5	10.7	1.8
PhD	-59.1	-76.7	12.2	3.5
Economy and administration				
Uncompleted high school	-17.4	-24.1	4.2	1.9
Completed high school	-28.4	-37.9	5.6	0.9
High school w/ext.	-34.9	-48.1	6.7	2.7
College, lower level	-43.4	-57.5	7.5	-3.3
College/University, higher level	-48.2	-66.3	9.4	-1.1
PhD	-60.0	-76.0	1.8	-7.3
Teacher education (College level)	-20.3	-33.9	4.9	10.1

The type of job also has a large impact on absenteeism, and entry rates typically vary by up to around 30 percent across major industries. Absence rates are highest in the manufacturing, teaching, and health care sectors, and lowest in the oil industry, retailing, and research and development. Absenteeism generally rises with work-hours and declines with earnings (given the number of work-hours). Within the group of full-time workers, we find that members of the upper earnings quartile have a 42 percent lower entry rate to major absence – and a 47 percent higher recovery rate – than members of the first earnings quartile, ceteris paribus. We also find that absenteeism declines strongly with wealth, which probably reflects that wealth is correlated with unobserved human capital characteristics as well as with social background.

Previous evidence has indicated that insecure jobs encourage workers to avoid absenteeism; see Arai and Thoursie (2005) and Ichino and Riphahn (2005). We have examined this hypothesis by using very short tenure (less than one year) as a proxy for job insecurity (most jobs in Norway come with a trial period of six or 12 months). However, we find no consistent evidence that short-tenured workers have less certified absence than more secure workers. Short tenure is associated with a 2.9 percent *higher* entry rate into minor diseases and a 5.7 percent lower hazard rate into major diseases.[10]

1.7 The workplace

Small workplaces (less than 20 employees) have entry rates into certified absence that are 20-25 percent lower than those of large workplaces, ceteris paribus. A possible interpretation of this finding is that workers at small workplaces to a larger extent than workers at large workplaces internalize the adverse consequences of own absence. Moreover, small workplaces are more transparent; hence it is more difficult to be absent without anyone noticing. The characteristics of colleagues have a significant impact on each employee's absence propensity. In particular we find that entry rates into both types of absence spells decline strongly with the average age of the colleagues at a workplace. The entry rates at workplaces with mean age above 50 years are around 8-12 percent lower than at workplaces with mean age below 40 years, ceteris paribus. Entry rates also decline with the average education level. These findings are all consistent with the idea that absenteeism is affected by local social norms; i.e., when the colleagues have characteristics implying low average absence propensity at the workplace, this also implies that each individual's threshold for claiming sick is higher than it would have been under other circumstances. This conclusion does not hold for the gender composition, however. We find that a balanced gender mix implies lower entry rates into major diseases than either male-domination or female-domination. Domination by one gender is also associated with low reentry rates from both types of absence. A possible interpretation of these findings is that a balanced gender composition is conducive to the work environment.

Workplaces with high employee turnover have approximately 6 percent higher entry rates into major disease absence and 2-3 percent lower recovery rates from both diag-

[10] We also exploit regional idiosyncrasies in labor market tightness to investigate the impact of job insecurity. Previous Norwegian evidence indicates that absenteeism vary procyclically because the threat of being laid off is more frightening the poorer are the chances of getting a new job (Askildsen, Bratberg, and Nilsen, 2005; Nordberg and Røed, 2009). Labor market tightness is represented in our model by the transition rate out of registered unemployment in the municipality. As it turns out, our findings confirm that there is a positive impact of labor market tightness on entry into minor diseases; a 10 percentage point increase in the probability of escaping unemployment raises the entry rate to minor absence by 3.3 percent. For the other transitions, we only find small or statistically insignificant impacts.

noses than workplaces with low turnover.[11] Downsizing reduces entry into minor disease absence (with up to 3.5 percent), but has no effect on entry into major disease absence. Downsizing also significantly reduces recovery rates, particularly from major disease absence (with up to 5.5 percent). The latter finding may reflect that employers going through downsizing have weak incentives to encourage recovery, given that the social security system pays the whole wage bill for long-term absentees.

1.8 The panel doctor

Table 4 reports the impacts of various panel doctor characteristics. A point to note is that workers listed with very young panel doctors (below 30 years) are significantly less absent from work than workers listed with older panel doctors. Recovery hazards also decline monotonously with the doctor's age. Female doctors seem to be "stricter" than male doctors. Having a female panel doctor reduces the entry rates into absence by around 2.5 percent, ceteris paribus. The effect of the doctor's gender is slightly larger for female than for male workers. Doctors sharing office with other doctors, e.g. in a medical centre, are stricter than doctors operating alone. Specialists are slightly stricter than non-specialists.

Table 4. The impacts of panel doctor characteristics (percentage change in hazard rates)

	Absence entry		Recovery	
	To minor	To major	From minor	From major
A. The panel doctor's age (reference=40-50 years)				
< 30 years	-3.7	-2.6	3.5	5.4
30 – 40 years	-1.1	-0.1	3.8	2.3
40 – 50 years	Ref.	Ref.	Ref.	Ref.
50 – 60 years	-0.6	-1.4	-2.9	-1.2
> 60 years	-0.2	-3.0	-7.6	-1.7
B. Female panel doctor (reference=male panel doctor)				
Male patient	-2.4	-2.3	0.5	-0.9
Female patient	-2.6	-3.3	0.8	-1.3
C. Panel doctor sharing office with other doctors (reference=operating alone)	-2.9	-1.4	0.8	0.4
D. Panel doctor has a specialist education (reference=non-specialist)	-0.4	-0.5	3.8	2.4

We have also investigated the impacts of the panel doctor's workload and competitive situation (the degree of patient shortage). Our findings indicate that doctors with

[11] The turnover rate is defined on a quarterly basis as Min(number of entries, number of exits) divided by the number of employees at the start of the quarter.

few patients, in general, and with fewer numbers of patients than desired, in particular, certify less absence than doctors with many patients. The competitive situation in the municipality – as measured by the panel doctors' average shortage of patients in the municipality – seems to be of minor importance. These findings may reflect that panel doctor assignment is selective. Patients are free to choose between doctors with spare capacity, and they may prefer lenient to strict doctors. On the other hand, if doctors respond strongly to financial incentives, we would expect panel doctors with few patients to be particularly lenient. Our results are more in line with the selection-of-doctors hypothesis than with the doctor-respond-to-incentives hypothesis.

1.9 Family events

We now turn to the impacts of important family events, such as pregnancies, separations/divorces, and the death of a close family member (spouse, child, mother, or father). All the impacts are estimated dynamically, i.e., we investigate the *time profile* of impacts on the hazard rates prior to and after the event in question actually occurs. These profiles are estimated non-parametrically, using time varying indicator variables that capture the effect of the time distance to the event in question (measured in weeks or months before/after). Figure 7 presents the estimated impacts on entry into the major diagnosis (it turned out that the events in question only had minor impacts on the other hazard rates).

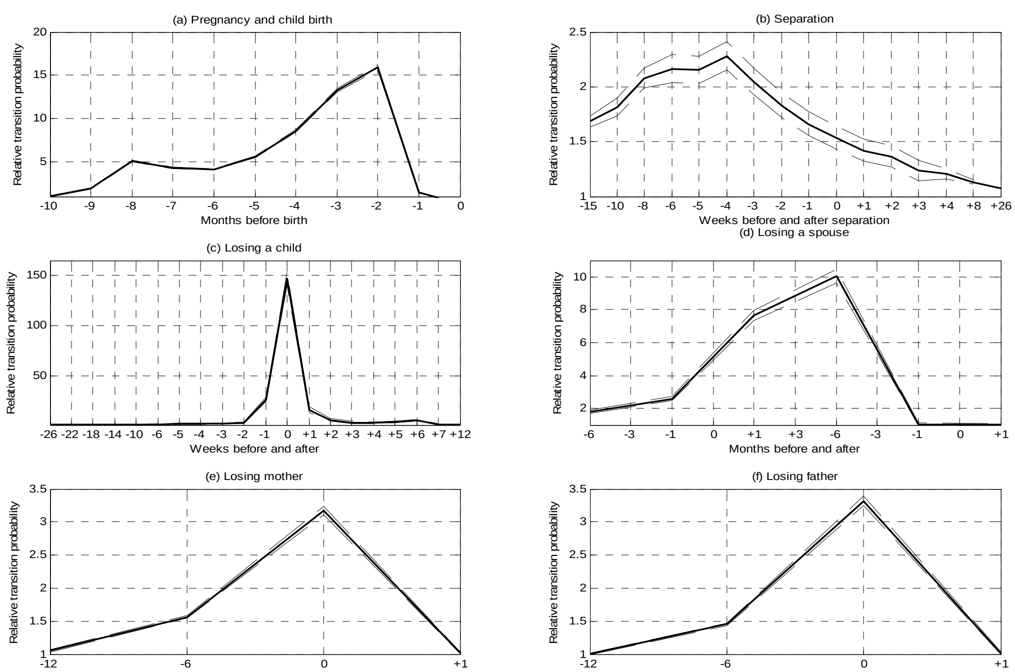

Figure 7. The estimated impacts of family shocks on the hazard rate to a major disease absence spell (with 95 percent confidence intervals; reference is no family shock).

Pregnancy has the almost immediate effect of five-doubling the rate of entry into major diseases absence spell; see panel a). The effect reaches a maximum two months before birth, at which point the entry rate is 15 times higher than before the pregnancy, ceteris paribus. The process of separation/divorce also entails increased risk of entry into major disease absence spell; see panel b). The impact is largest around 4-8 weeks prior to separation (at which point the entry rate is raised by a factor of 2.3), and it declines rapidly after the couple has actually split.

The loss of a close family member clearly has a huge impact of the entry into a major disease absence spell, see panels c)-f). In particular, the loss of a child or a spouse raises the hazard dramatically, by a factor of 150 and 10, respectively. But the loss of parents also entails significantly higher absence rates, particularly during the six months prior to the parent's death. This finding suggests that the demand for informal care may constitute a burden for employed offspring during the terminal phase of parents' lives, in line with recent findings reported by Fevang, Kverndokk, and Røed (2008).

The large impact of family events on absenteeism is consistent with the interview-based finding reported by Carlsen (2008) that panel doctors sometimes certify that employees are sick even when this is not strictly true. This may reflect that the sickness absence insurance system in Norway has developed into a more general "justified absence" insurance system, where physicians certify sickness to help employees cope with a difficult life situation.

1.10 Unobserved heterogeneity

Unobserved heterogeneity affects the four hazard rates through a joint discrete distribution of worker-specific intercepts, as reflected in the 29-support-points discrete distribution. By construction, unobserved covariates are orthogonal to the observed covariates. As we show in the next section, unobserved heterogeneity turns out to account for more of the cross-sectional variation in absence propensities than all the observed covariates taken together. Hence, it may be of interest to take a closer look at the structure of unobserved heterogeneity. Table 5 presents the estimated rank correlation (Kendall's τ) between the four unobserved intercepts.[12] There is a strong positive correlation in the minor (v_{12}) and major (v_{13}) absence entrance propensities. There is apparently no systematic direct relationship between unobserved entry and recovery propensities. We find, however, that workers with a high probability of entering minor absence (v_{12}) tend to have a high recovery rate from major absence (v_{31}), and that workers with high probability of entering major absence (v_{13}) have a low recovery rate from minor absence (v_{31}).

Table 5: Estimated rank correlation between the four unobserved covariates (Kendall's τ)

	v_{12}	v_{13}	v_{21}	v_{31}
v_{12}	1	0.678	-0.055	0.195
v_{13}		1	-0.272	0.018
v_{21}			1	0.645

Note: Rank correlation is calculated from the 29-points discrete distribution. See footnote for details.

5. The cross-sectional variation in steady state absence rates

The purpose of this section is to decompose the variation in absenteeism *between employees* into its appropriate observed and unobserved determinants. This cannot be done on the basis of recorded absence at a particular point in time, since each employee is then either absent or present, and little information is revealed regarding the underlying absence propensities. We should therefore clearly examine the absence behavior over some period of time. As the time-period becomes longer, the random nature of health shocks becomes less important, and each employee approaches an average absence rate corresponding to his/her intrinsic absence propensity. The influence of window length on the observed variance in mean individual absence rates is illustrated in the two upper panels of Figure 8. As the time window is extended, the cross sectional variance in absence rates declines. We observe the same pattern whether we look at actual absence behavior or at the absence behavior *simulated on the basis of our statistical model*, although the simulated data clearly underrates the cross-sectional variance of major disease absence somewhat.

[12] Kendall's τ is computed on the basis of all possible pairs of workers *(i,j)* that can be formed from the heterogeneity distribution. A pair $\left\{(v_{ki}, v_{li}), (v_{kj}, v_{lj})\right\}$, $k,l = 12, 13, 21, 31,$ is concordant with respect to variables *(k,l)* if $(v_{ki} - v_{kj})(v_{li} - v_{lj}) > 0$ and discordant if $(v_{ki} - v_{kj})(v_{li} - v_{lj}) < 0$. Let c_{kl} be the number of concordant pairs and let d_{kl} be the number of discordant pairs. We then compute Kendall's τ as $\tau_{kl} = \dfrac{c-d}{c+d}$. We disregard the fraction $\sum_{s=1}^{Q} p_s{}^2$ of pairs drawn from the same location vector.

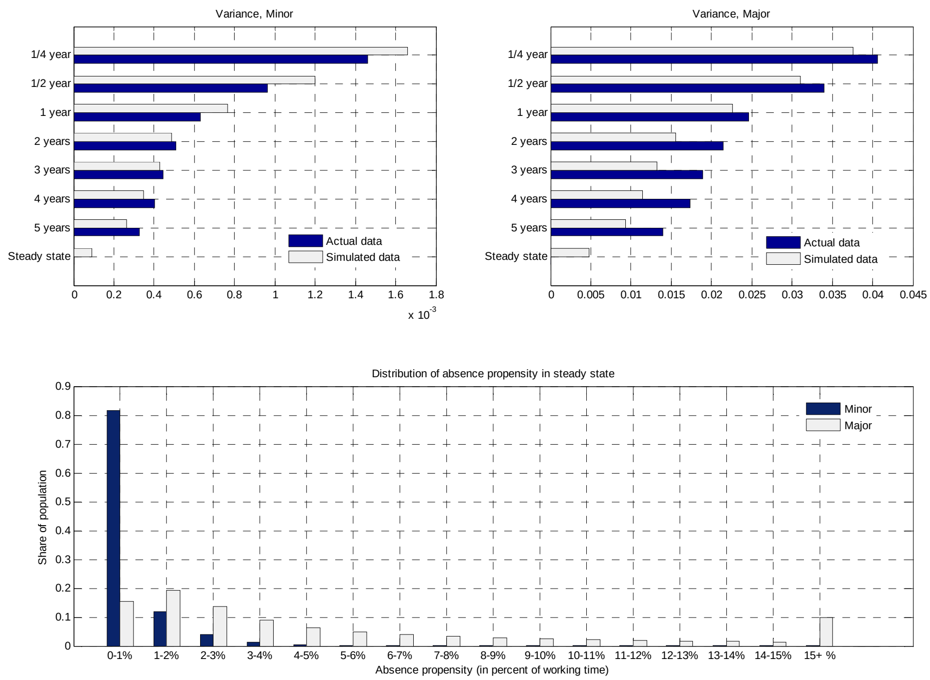

Figure 8. Variation in individual absence rates.
Note: The upper panels illustrate the distribution of mean individual absence rates by length of the time window considered. The lower panel shows the distribution of steady state absence rates.

1.11 Steady state absence rates

Based on our estimated model, we can predict each employee's long-run absence propensity, defined formally as the absence rate that prevails as the time window goes towards infinity. We compute these *steady state absence rates* 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 that recovery probabilities vary with absence duration.[13] 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 spent in sickness absence over an infinitely long time horizon, given that no changes occur in individual characteristics or in environmental factors. The steady state absence rates examined below are computed from the predicted hazard rates in the third quarter of 2002, but the cross-sectional decomposition is not at all sensitive

[13] From the covariates and the estimated coefficients we construct one transition matrix for each individual in the sample. The limiting distribution then gives us the expected share of a persons time spent in each state. Formally, there are more than 100 states in our transition matrix, since recovery rates are allowed to vary by duration.

with respect to this particular selection of point in time. Looking at the bottom bars in the upper panels of figure 8, we conclude that the steady state variance is roughly half of the five-year variance for both minor and major absence. The reported variances imply that the standard deviations in steady state absence rates are around 8 percent for major absence and 1 percent for minor absence. The mean (median) steady state absence rates are 0.6 (0.3) percent for minor disease diagnoses and 5.9 (3.2) percent for major disease diagnoses. The lower panel of Figure 8 describes the distributions of steady state absence rates in more detail. While around 80 percent of the employees are absent less than 1 percent of their potential working days due to minor diseases, the major disease absence propensity is subject to large variation, with as much as 10 percent of the employees having steady state absence rates above 15 percent.

1.12 Variance decomposition of steady state absence rates

We now turn to the issue of decomposing the variation in absence rates that our model does account for into its various sources, and to disentangle direct impacts of explanatory variables from impacts operating through correlation with other variables. Let $\hat{a}_{ki}$ be individual *i's* predicted steady state absence rate corresponding to state *k* (*k=2,3*) (minor, major) We suppress the time-subscript here to emphasize the steady state nature of the calculations; all time-varying covariates are held fixed. Let $z_i = \{x_i, v_i\}$ be the complete set of explanatory variables determining $\hat{a}_{ki}$, and let z_{ci} be a subset of these variables. We can now investigate how much of the variance in $\hat{a}_{ki}$ which is accounted for by the variable subset z_c based on the law of total variance, i.e.,

$$\operatorname{var} \hat{a}_{ki} = \underset{c}{E}\left[\operatorname{var} \hat{a}_{ki} \mid z_c\right] + \underset{c}{\operatorname{var}}\left[E\left[\hat{a}_{ki} \mid z_c\right]\right], \qquad (5)$$

The fraction of total variance accounted for by z_c is thus $\left[\underset{c}{\operatorname{var}}\left(E\left[\hat{a}_{ki} \mid z_c\right]\right)\left(\operatorname{var} \hat{a}_{ki}\right)^{-1}\right]$. Part A of Table 6 reports the percent of overall variance across individuals – in total, minor, and major absence – that is attributable to the variance between the different categories that can be established on the basis of various explanatory variable groups. The number of categories varies a lot across the different variable groups. For example, for the variable group "age", we simply divide the population into 31 groups depending on age measured in years. For the variable group family background, on the other hand, we construct 280 categories based on all possible combinations of the different family background covariates. And for workplaces and panel doctors, we let each workplace and doctor constitute a separate category. The results suggest that job characteristics such as wages, work-hours and seniority account for 11.5 percent of the variance in expected steady state absence rates, while the identity of the workplace accounts for 7.9 and the worker's educational attainment for 7.4 percent. The most important factor, however, is unobserved heterogeneity, which according to our model accounts for as much as 66 percent of the variation in steady state absence rates.

Table 6: Percent of cross-sectional variance in steady state absence rates attributable to different variable groups

Variable group (number of categories included within each variable group)	*A. Gross variance decomposition: Percent of total variance accounted for by variable group*			*B. Partial variance decomposition: Percent of total variance accounted for by variable group*			*C. Partial variance decomposition: Percent of total variance accounted for by fixed effects**		
	Total	*Minor*	*Major*	*Total*	*Minor*	*Major*	*Total*	*Minor*	*Major*
Age (31)	0.8	1.2	1.1	1.1	1.1	1.4			
Family background (280)	2.4	1.0	2.4	0.6	0.3	0.6			
Gender and family situation (13)	5.2	2.5	5.2	2.2	2.3	1.9			
Education (65)	7.4	1.7	7.5	2.9	1.8	2.7			
Place of residence (99)	1.2	0.9	1.4	0.7	1.0	0.7			
Job characteristics (wage, hours, seniority) (25)	11.5	1.8	12.1	9.1	0.9	9.7			
Local labor market (10)	0.1	0.1	0.1	0.0	0.0	0.0			
Country of origin (3)	0.2	0.7	0.2	0.1	0.3	0.1			
Workplaces (3820)	7.9	3.0	8.0	0.7	0.9	0.7	3.5	3.3	3.6
Panel doctor (3522)	2.1	1.4	2.2	0.0	0.1	0.0	2.0	2.4	1.9
Unobserved heterogeneity (29)	66.0	76.1	64.8	66.1	81.4	64.5			

*Based on results from a separate model with fixed workplace and panel doctor effects; se text.

These fractions clearly reflect a combination of sorting on the conditioning variables and their possible causal effects on absence behavior. For example, the fraction of absence variance accounted for by workplaces partly reflects that different workplaces recruit different workers and partly that the absence behavior of each worker is affected by workplace characteristics. Our multivariate model gives us the opportunity to eliminate sorting attributable to other factors, and hence to assess the partial influence of particular (groups of) variables. Assume that we wish to assess the partial influence of a subset of observed variables x_c. Let x_{-c} denote the remaining observed characteristics. We then write the four predicted hazards as $\hat{\theta}_{ki}(x_{ci}, x_{-ci}, v_{ki}) = \exp(x_{ci}\hat{\beta}_{jkc})\exp(x_{-ci}\hat{\beta}_{jk-c} + v_{ki})$. To "remove" all sources of variation other than those caused by x_c, we compute steady state absence rates for all employees with the proportionality factors $\exp(x_{-ci}\hat{\beta}_{jk-c} + v_{ki})$ replaced by their respective means, i.e., $\frac{1}{N}\sum_{i=1}^{N}\exp(x_{-ci}\hat{\beta}_{jk-c} + v_{ki})$.

This implies that we compute the steady state absence rate as a function of x_c for workers who are representative along other dimensions. Let $\hat{a}_{ki} \mid z_c, \overline{z}_{-c}$ be worker i's steady state absence rate predicted this way. We then compute the partial variance contribution from variable group z_c as $\left[\underset{c}{\text{var}}\left(E\left[\hat{a}_{ki} \mid z_c, \overline{z}_{-c} \right] \right) \left(\text{var}\, \hat{a}_{ki} \right)^{-1} \right]$.

Part B of Table 6 reports the results from the partial variance decomposition, i.e., the fraction of variance explained by a given variable group when all other variables are held fixed at a "representative" level. We first note that the fraction of the variance explained by unobserved characteristics remains virtually unchanged. This is exactly what we would expect, since unobserved heterogeneity by construction is distributed independently of all other covariates; hence there is no systematic sorting on unobserved heterogeneity.[14] For the other covariate groups, however, the partial influence is typically much smaller than indicated by their gross influence. The only observed variable group still accounting for a major part of the variance is the one reflecting job characteristics (work-hours and wages). Workplaces and panel doctors no longer seem to matter very much. There are two possible interpretations of this latter finding. The first is that workplaces and doctors have little influence on absence behavior and that the large variation in absence rates across workplaces and panel doctors solely results from employee sorting. The second interpretation is that observed workplace and doctor covariates (see the Appendix) do not appropriately represent the true variation in workplace characteristics and doctor practices.

1.13 Variance decomposition based on a fixed effects model

To assess this latter hypothesis, we estimated an additional version of the hazard rate model outlined in Section 3, this time with all observed workplace and panel doctor characteristics replaced by workplace and panel doctor dummy variables. To make the estimation of this model feasible and to avoid too much noise in the estimated workplace and panel doctor fixed effects, we dropped all employees in workplaces with less than 100 employees or with panel doctors with less than 100 employed patients. This model nevertheless implied estimation of as much as 35,000 parameters. Given the computational challenge, we were not able to include unobserved heterogeneity in this model. The estimated parameters from the fixed effects model were in turn used to compute a new set of

[14] Note, however, that this independence result will not hold exactly, since the relationship between explanatory variables and steady state absence rate is non-linear.

steady state absence rates. And again, we calculated the variance in representative employees' predicted steady state absence rates, conditional on workplace and panel doctor, respectively (as explained in the previous subsection). The results are reported in part C of Table 6. Note that we relate the variance generated by workplaces and panel doctors in the fixed effects model with the total variance generated from the model with unobserved heterogeneity included (based on the full population), but without fixed effects; hence this exercise does not constitute a valid variance decomposition. The reason why we do it this way is that the overall variance in steady state absence rates drops enormously when we do not account for unobserved heterogeneity, hence the "explained part" rises accordingly for all variable groups. To make the numbers in Table 6 directly comparable, we find it most illuminating to relate all the between-group variances to the same total variance.

The results from the fixed effects model show that there is a significant variation in steady state absence rates across workplaces and that this variation cannot fully be accounted for by sorting. Yet, our workplace dummy variables do not explain more than around 3.5 percent of the overall variance in steady state absence rates, while panel doctor dummies explain around 2.0 percent. A comparison of these numbers with the gross variance contributions reported in part A of the table indicates that while sorting accounts for roughly half of the variance in absence rates across workplaces, it accounts for virtually nothing of the variance across panel doctors.[15]

1.14 The roles of workplaces and physicians revisited

Although workplaces and physicians are responsible for a modest part of the overall variance in steady state absence rates, their behavior does have a significant impact on absenteeism. Based on estimates from the fixed effects model, Figure 9 shows how the long-term absence rate of a representative worker changes as he/she moves from the workplaces and doctors associated with lowest absence rates towards the workplaces and doctors associated with highest absence rates. These partial – and potentially causal – effects are compared with the actual distribution of five-year average absence rates observed in the data (reproduced from Figure 1). Since sampling-error alone clearly generates a purely spurious difference between workplaces and panel doctors in the fixed-effects

[15] Note, however, that by focusing on large workplaces only (more than 100 employees), we disregard an important source of workplace heterogeneity, since the analysis in Section 4.4 revealed substantial differences between large and small workplaces.

model, we also provide two “placebo” distribution curves. These have been generated by re-estimating the fixed effects model with counterfactual random-assignment of workplaces and panel doctors; i.e., we have constructed artificial workplace and panel doctor indicator variables that by definition have no effect on absenteeism.

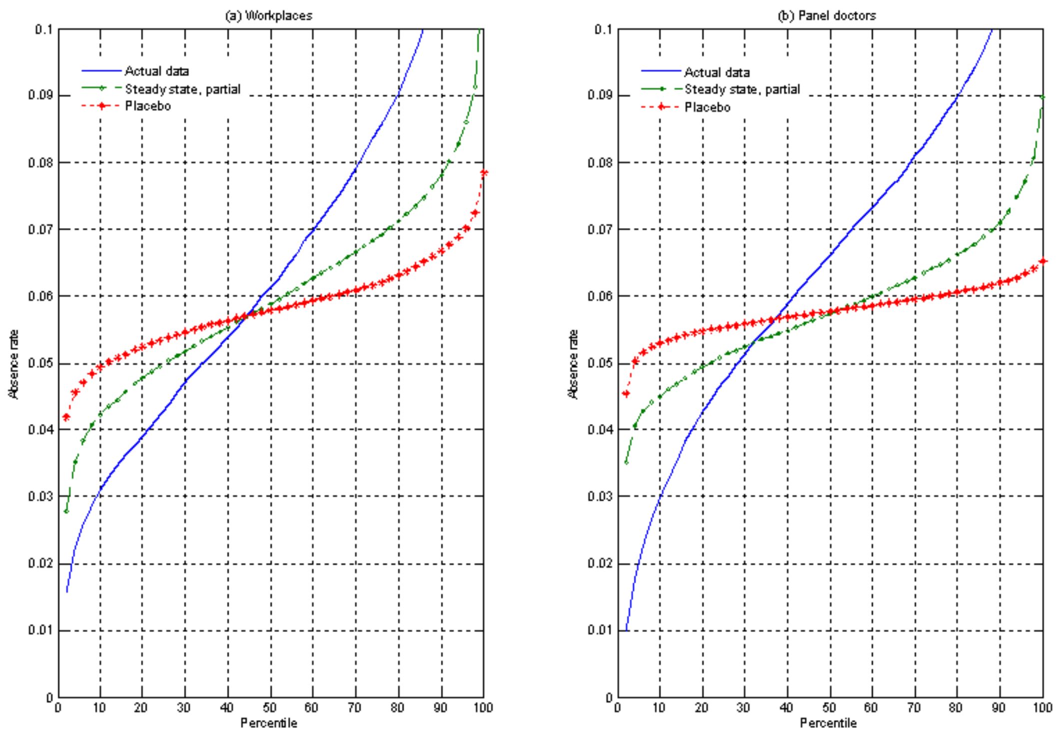

Figure 9. The distribution of absence rates across workplaces and panel doctors.
Note: The graphs are based on actual five-year averages and on steady state absence rates calculated from a fixed effects model (without unobserved heterogeneity).

It is clear from these graphs that even though a large part of the absence variation across firms is caused by sorting (and arbitrariness resulting from the limited time-window available for comparison of actual absence rates), a significant workplace-effect remains when all other variables are controlled for. Moving from the 10th to the 90th percentile in the estimated workplace effect distribution raises steady state absence rate from around 4.2 to 7.8 percent, ceteris paribus; i.e., by 86 percent. In the corresponding placebo estimation, the difference is 35 percent. Hence, we conclude that most of the estimated variation across workplaces is really causal.

Panel doctors also seem to have a significant impact on their patients’ absence behavior. Moving from the 10th to the 90th percentile in the panel doctor effect 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 panel doctors seems to be causal.

6. Discussion and conclusion

The results presented in this paper show that the vast majority of employees (80 percent) on average will be absent (certified by a doctor) less than 1 percent of their working days due to minor – and easily verifiable – “everyday”-diseases, such as respiratory infections and virus diseases. The contribution made by this type of absence to the overall variance in absence rates is minimal. By contrast, absence due to potentially more serious – and less verifiable – diseases, such as musculoskeletal and mental disorders, are extremely unequally distributed across employees, and as much as10 percent of the workforce can expect to be absent more than 15 percent of the time. Most of this variation remains “unexplained” in our statistical analysis, however, despite our access to an extraordinary rich set of explanatory variables covering employees, their workplaces, their doctors, and the economic environments they operate in. Yet, we do identify empirical regularities that shed light on the anatomy absenteeism, and on the roles that workers, firms, doctors, and certification regulations, may have in bringing it down.

Although most of the variation in absenteeism across workplaces reflects employee sorting, large differences remain when we control for sorting and random variation. These differences can only to a limited extent be explained by observed workplace characteristics. We find, however, that each worker’s absence behavior is affected by the characteristics of colleagues, suggesting that the development of workplace-specific social norms is a part of the story.

Physicians’ absence certification practices also have a significant impact on patients’ absence behavior, and simply switching to a new panel doctor may cause large shift in an employee’s absence propensity. This clearly illustrates the large scope for judgment regarding the use of *rest* as an appropriate treatment for many long-term ills, like musculoskeletal and mental diseases. Again, we have difficulties identifying observed doctor-characteristics that can explain the variation in certification practices. We do find, however, that strict doctors tend to have few patients, and often fewer than they actually desire. We also find that strictness declines with the doctor’s age, and that female doctors are stricter than male doctors. We interpret the estimated calendar time profiles in the sickness absence entry and recovery hazards as convincing evidence that the

2004-reform in certification regulations – with larger emphasis on activation requirements during sickness absence periods – did cause a significant drop in absenteeism.

A somewhat surprising result is that the frequency of both types of absence spells *decline* strongly with age up to around 45 years, even when we control for all the factors typically assumed to cause higher absence among younger employees, such as responsibility for own children and care-needing parents. We also find that each employee's absence propensity declines with the average age of the colleagues at the workplace. Taken together, these findings suggest that younger employees have lower thresholds for claiming sick than older employees, ceteris paribus. This may either indicate that social norms regulating the exploitation of sickness absence insurance are weaker among young than old workers, or that such norms are about to deteriorate. We are also surprised by the large difference in absenteeism between men and women that remains after we have controlled for factors such as children, pregnancies, education, occupation, and pay. Depending on family situation and type of sickness, females' entry rates to certified sickness absence spells are between 33 and 75 percent higher than those of similar males.

In line with existing evidence we identify a strong "social gradient" in sickness absence behavior. The gradient prevails regardless of whether we measure status by family background, own educational attainment, occupation, wealth, or pay. Higher social status is *always* associated with lower sickness absence. A particularly interesting feature of the relationship between education and absenteeism is that the *level* of education seems to be much more important than its type, suggesting that there is a general "education effect" operating over and above the type of job that the education qualifies for.

We find strong evidence that the sickness insurance system is exploited extensively to offer employees a respite from work in relation to traumatic personal events, such as marriage dissolution, seriously ill (dying) parents, and the loss of close family members. This may to some extent reflect that these events actually *cause* sickness. But it is also likely to reflect that sickness certification is used as a sort of social insurance of last resort. Sickness absence also rises dramatically during pregnancies. Taken together, these findings suggest that the sickness insurance system effectively insures employees against a broad range of circumstances that make it difficult to go to work, despite that the legislation explicitly states that it covers own *diagnosed sickness* only.

Given that the earnings of virtually all employees in Norway are 100 percent insured against sickness absence for up to one year, there is limited scope for assessing absence responsiveness with respect to financial incentives. After one year of absence, however, those who do not return to work are transferred to the less generous "rehabilitation benefit", which on average is associated with an income drop around 35 percent. We find a huge rise in the recovery probability at this point; for a typical employee, the weakly recovery probability rises from around 5 to almost 30 percent during the very last weeks of the sickness insurance period. This response clearly indicates that there is no such thing as a deterministic relationship between health and absence. Insurance institutions and panel doctor behavior matter. And so do the economic incentives of the employee.

Despite that we have exploited an extraordinary rich dataset – with almost 400 explanatory variables characterizing employees, workplaces, doctors, and economic environments – most of the cross-sectional variation in absenteeism is ultimately "explained" by unobserved time-invariant individual characteristics. To the extent that these characteristics are not only unobserved by us (the researchers), but also by economic agents, their explanatory power provides a rationale for statistical discrimination in hiring policies, whereby employers use observed sickness absence behavior as a screening device; see, e.g,. Amilon and Wallette (2009).

References

Amilon, A, and Wallette, M. (2009) Work Absence - A Signalling Factor for Temporary Workers? *Labour*, Vol. 23, No. 1, 171–194.

Arai, M. and Thoursie, P. S. (2005) Incentives and selection in cyclical absenteeism, *Labour Economics*, Vol. 12, No. 2, 269-280.

Askildsen, J. E., Bratberg, E., and Nilsen, Ø. A. (2005) Unemployment, Labour Force Composition and Sickness Absence. A Panel Data Study. Health Economics Vol. 14, No. 11, 1087-1101.

Bonato, L. and Lusinyan, L. (2004) Work Absence in Europe. IMF Working Paper 04/193 (www.imf.org/external/pubs/ft/wp/2004/wp04193.pdf)

Card, D., Chetty, R. and Weber, A. (2007) The Spike at Benefit Exhaustion: Leaving the Unemployment System or Starting a New Job? *American Economic Review*, Vol. 97, No. 2 (PAP), 113-118.

Carlsen, B. (2008) Dobbeltmoralens voktere? Intervjuer med fastleger om sykemelding. *Tidsskrift for Velferdsforskning*, Vol. 11, No. 4, 259-275.

Fevang, E., Kverndokk, S, and Røed, K. (2008) Informal Care and Labor Supply. IZA

Discussion Paper No. 3717.

Gaure, S., Røed, K. and Zhang, T. (2007) Time and Causality – A Monte Carlo Evaluation of the Timing-of-Events Approach. *Journal of Econometrics*, Vol. 141, 1159–1195.

Goffe, W.L., G.D. Ferrier, and Rogers, J. (1994) Global Optimization of Statistical Functions with Simulated Annealing. *Journal of Econometrics*, Vol. 60, 65-99.

Heckman, J. and Singer, B. (1984) The Identifiability of the Proportional Hazard Model. *Review of Economic Studies*, Vol. 51, 231-241.

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

Ichino, A. and Riphahn, R.T. (2005) The Effect of Employment Protection on Worker Effort: Absenteeism During and After Probation. *Journal of the European Economic Association*, Vol. 3, No. 1, 120-143.

Ichino, A. and Moretti, E. (2009) Biological Gender Differences, Absenteeism and the Earning Gap, *American Economic Journal: Applied Economics*, Vol. 1, No. 1, 183-218.

Johansson, P. and Palme, M. (2002) Assessing the Effect of Public Policy on Worker Absenteeism. *Journal of Human Resources*, Vol. 37, No. 2, 381-409.

Lindsay, B. G. (1983) The Geometry of Mixture Likelihoods: A General Theory. *The Annals of Statistics*, Vol. 11, 86-94.

Markussen, S. (2009) Closing the gates? Evidence from a natural experiment on physicians' sickness certification. In progress.

Marmot, M. (2004) *The Status Syndrome. How Social Standing Affects Our Health and Longevity*. Time Books, New York.

Martocchio, J. J. (1989) Age-related Differences in Employee Absenteeism: A Meta-Analysis. *Psychology and Aging*, Vol. 4, No. 4, 409-414.

Nordberg, M. and Røed, K. (2009) Economic Incentives, Business Cycles, and Long-Term Sickness Absence. *Industrial Relations*, Vol. 48, No. 2, 203-230.

Ruhm, C. J. (2000) Are Recessions Good for you Health? *Quarterly Journal of Economics*, Vol. 115, 617-650.

Røed, K. and Zhang, T. (2003) Does Unemployment Compensation Affect Unemployment Duration? *The Economic Journal*, Vol. 113, 190-206.

Taylor, H. M. and Karlin. S. (1998): *An introduction to stochastic modelling*, Academic Press

Appendix

Table A1. Overview of explanatory variables

I. Demographic factors	Gender (2 dummy variables, interacted with marital status and presence of children below 12 years of age) Age (31 dummy variables) Nationality (3 dummy variables) County (20 dummy variables) Size of municipality (5 dummy variables)
II. Family background	Parents' education (8 dummy variables) Parents' health – indicators for early disability or death (6 dummy variables) Parents' income (6 dummy variables)
III. Current family situation	Marital status (3 dummy variables, interacted with gender and presence of children below 12 years of age) Earnings relative to the spouse (3 dummy variables, interacted with gender) Dependent children (2 dummy variables, interacted with gender and marital status)
IV. Family events/shocks	Death in close family (36 dummy variables, covering periods prior to and after decease of spouse, parent or child) Divorce (17 dummy variables) Pregnancy (17 dummy variables)
V. Own human capital and job characteristics	Length and type of own education; (65 dummy variables) Earnings and work hours (13 dummy variables) Wealth (5 dummy variables)
VI. Job security	Tenure (2 dummy variables) Local labor market tightness (1 scalar variable)
VII. The workplace	Industry/sector (62 dummy variables) Size (2 dummy variables) Turnover (6 dummy variables) Ongoing and completed downsizing/upsizing (12 dummy variables) Age composition of workforce (6 dummy variables) Gender composition of workforce (6 dummy variables) Education composition of workforce (6 dummy variables) Participation in IWA, private and public sector (4 dummy variables)
VIII. Sorting of the workforce	Local employment rate (1 scalar variable)
IX. The panel doctor	Number of patients (8 dummy variables) Shortage of patients in municipality (4 dummy variables) Deviations from the desired number of patients (6 dummy variables) Specialty (2 dummy variables) Co-practice (2 dummy variables) Age (5 dummy variables) Gender (4 dummy variables, interacted with gender of patient) Numbers of doctors per capita in municipality (6 dummy variables)
X. Time	Quarter (19 dummy variables)
XI Spell duration	Piece-vice constant baseline hazard (28 duration dummy variables)

Note: The reported dummy variable numbers include the reference categories.