Draft. Do not quote.

26 May 2008

The Anatomy of Absenteeism

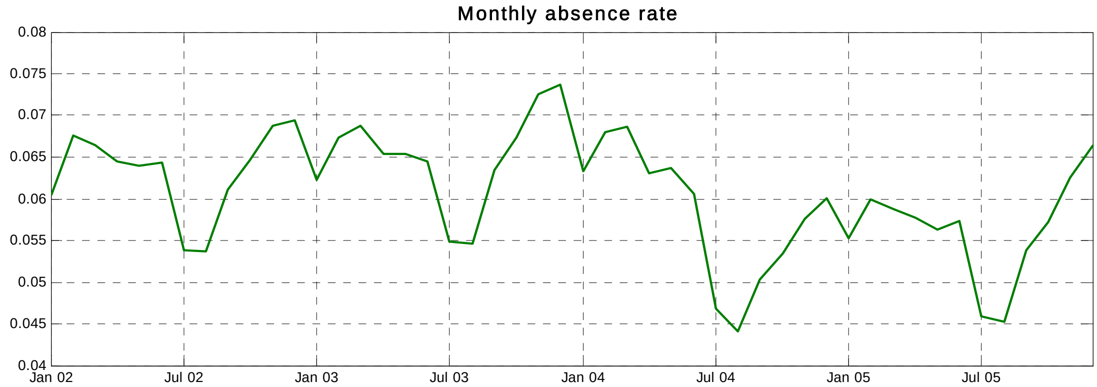
Simen Gaure, Simen Markussen, Knut Røed, Ole J. Røgeberg

1. Introduction

On a typical working day, 6-7 per cent of Norwegian employees are absent from work due to sickness. In total, sickness absence insurance payments amount to approximately 41 billion NOK per year (2004), or around 2.4 per cent of GDP. Hence, the costs associated with sickness absence – in terms of forgone labor supply as well as direct insurance payments – are substantial. It is a well known fact that the level of sickness absence displays large variation across both time and countries; see, e.g. Bonato and Lusinyan (2004). But the behavioral mechanisms responsible for this variation are not well understood. Norway is among the countries in Europe with the highest rates of sickness absence, and apart from a brief period following an institutional reform in 2004 absenteeism has displayed an increasing trend since the early 1990's. In order to explain trends and disparities in absenteeism, economists tend to focus on the design of sickness insurance systems, employment protection legislation and cyclical fluctuations; see, e.g., Henrekson and Persson (2004), Ichino and Riphan (2005), Ruhm (2000), and Johansson and Palme (2002) for recent empirical evidence. But, although the causal impact of labor market institutions and economic fluctuations have been convincingly demonstrated in these as well as a number of other papers, their substantive significance remains disputed. While sickness absence is high on the political agenda in Norway, the attention has clearly turned away from individual incentives and absence behavior and towards the roles played by firms and physicians. This has, e.g., materialized in a tripartite "inclusive workplace agreement" between employers, employees and the state, devising strategies to improve work environments in general and to reintegrate absent employees in particular. Moreover, the guidelines regulating the physicians' sickness absence certification practice have been revised in order to encourage work presence rather than absence for employees with chronic health problems.

This paper examines the anatomy of sickness absence, based on complete register data for all certified absence spells in Norway – their starting dates, their stopping dates, and their outcomes – from 2001 to 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.

This paper explores the sources of variation in sickness absence propensity; in terms of individual factors (gender, age, work-experience, tenure, nationality, health, educational attainment, income, occupation, social background, family events, etc.), in terms of workplace-characteristics (industry, employee-composition, organizational changes), in terms of the local economic environment (job-finding rate, employment rate), in terms of panel doctor characteristics (age, gender, specialization), in terms of trends and fluctuations, and in terms of institutional characteristics (system reform). A key purpose is to decompose the variation across time, space, firms, GP's etc. into its appropriate sources. Figure 1 illustrates the degree of observed absence variation along some key dimensions. From the upper panel, we note that the average absence rate increased somewhat during the fall 2003. It then declined sharply until September 2004, after which it again started to rise. 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.



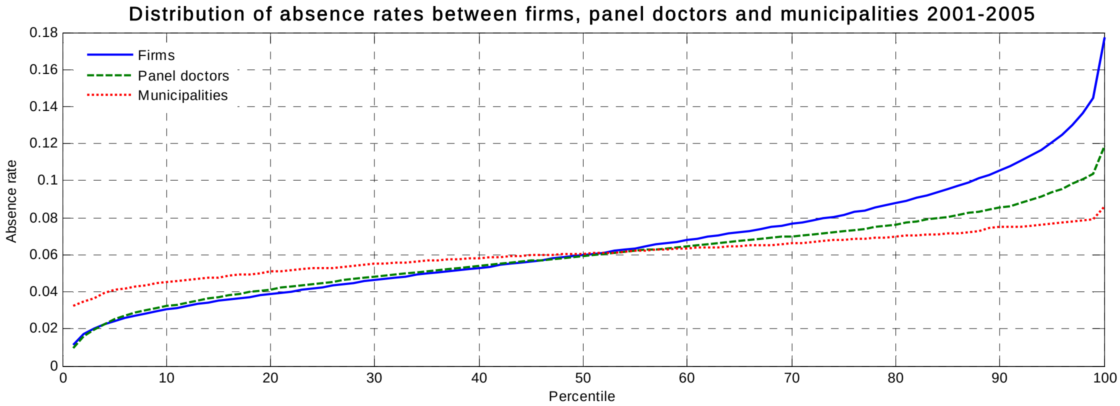


Figure 1: Absenteeism in Norway: The upper panel displays average absence rate 2002-2005. The lower panel displays the distribution of mean absence rates 2001-2005 within firms (>50 employees), GPs (>50 registered patients) and municipalities.

The lower panel illustrates the distributions of mean absence rates over all these five years by firms (with more than 50 workers), panel doctors (with more than 50 patients), and municipalities. The variation is substantial. For example, while firms in the 10^{th} percentile of the firms' absence rate distribution on average had absence rates around 3 percent, firms in the 90^{th} percentile had absence rate close to 11 percent. The corresponding numbers for panel doctors were 3 and 8 percent and for municipalities 4.5 and 7 percent. Absence variation across firms, panel doctors and municipalities reflects a combination of at least three types of factors: i) genuine randomness, ii) nonrandom sorting of employees (or other confounding factors), and iii) causal impacts of work environments, panel doctor practices, and local economic conditions, respectively.[1] The random component is of course more important the smaller are the entities under consideration, which can explain why the variation is larger among firms and doctors than among municipalities. A key purpose of this paper is to quantify the various contributions to absence variation, and to examine the underlying driving mechanisms.

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.

Our main findings are: TBW

2. Data and institutions

The data we use comprise starting dates and stopping dates for all certified sickness absence spells in Norway from June 2001 to December 2005. 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

[1] Given firms, doctors and municipalities are geographically tied to each other, all three factors generally affect all the distributions displayed in Figure 1.

bill.[2] The general rule is that absence spells 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”.[3] After 8 weeks of absence, a more thorough medical examination/certificate is required. In July 2004, a new regulation was implemented, implying that *partial* sickness absence should be the “default” option after this examination, i.e., the employee should start working *to some extent* unless there are strong medical reasons suggesting otherwise. This reform was implemented as a part of a general drive towards “activity orientation” of Norwegian social welfare policies.

Our data include information about the cause behind each absence spell (medical diagnosis/symptoms) and the (encrypted) identity of the physician responsible for its certification. The data also identify each citizen’s panel doctor. 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.

We merge the data on absence spells with a number of other administrative data registers containing information about individual employees, their panel doctors, their workplaces, and the institutional and economic environments they face. A summary of the resultant information regarding potential determinants of absenteeism is provided in Table 1.

---

[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.

Table 1.
Potential determinants of absenteeism and their representation in the data

I. Demographic factors	Gender Age Nationality
II. Family background	Parents' education Parents' health (indicators for early disability or death) Parents' income
III. Current family situation	Marital status Relative earnings (relative to the spouse) Dependent children Caretaking of own parents
V. Family events/shocks	Death in close family Divorce Pregnancy
VI. Own human capital and job characteristics	Educational attainment Earnings Wealth Working hours
VII. Absence incentives and transparency	Tenure (job security, risk of losing the job) Local labour market tightness (consequences if the job is lost) Size of municipality
VIII. The workplace	Size Member of "inclusive workplace agreement" Turnover Recent downsizing/upsizing Employee composition
IX. Sorting of the current workforce	Local employment rate
X. The panel doctor	Number of patients Shortage of patients in municipality Sufficiency of patients Speciality Co-practice Age Sex Doctors per capita in municipality
XI. Time	Institutional changes (July 2004) Important sports events Trends and cycles

Table 2 gives some key descriptive statistics regarding the population under study, their working environments, and the GPs they were attached to. We present descriptive statistics for the first and last year of our observation period, respectively, in order to be able to spot significant trends. We note, however, that the table does not reveal any conspicuous changes in the composition of the workforce during our data window. Average age increased slightly (by around half a year), the fraction of females increased (by one percentage point), and the fraction of persons with only compulsory education declined (by 2.7 percentage points).

Table 2. Key descriptive statistics

	2001	*2005*
The population		
Age (mean)	44.08	44.48
Percent females	47.04	48.16
Education		
Elementary school	10.90	8.19
High school	56.11	54.58
Higher education	31.74	35.14
Nationality (percent) Norway/OECD/World	90.44 / 7.02 / 2.49	89.04 / 7.72 / 3.12
Percent married or cohabitants	61.95	58.98
Percent with children < 12 years	41.41	39.38
The working environment		
Number of employees at the workplace (mean/median)	9.73 / 3	9.55 / 3
Mean turnover rate		
Percent of employees subject to downsizing	22.29	29.76
Percent with more than 1 year tenure in current job	62.98	62.18
Percent with full time job	79.50	79.61
Percent covered by the inclusive workplace agreement (IWA)	0	51.01
The economic environment		
Mean cyclical condition (the monthly transition rate from unemployment to employment for a “typical” unemployed)	0.17	0.14
The GPs		
Age (percent)		
below 40 years old	12.68	12.22
between 40-50 years old	23.27	18.69
above 50 years old	20.16	29.40
Education as specialist (percent)	32.83	26.16
Co-practice (Shares office with other GP)	4.74	5.82
Sufficiency of patients (percent)		
less than desired	38.93	49.16
more than desired	16.82	10.43
Gender (percent)		
Unknown GP characteristics	43.95	40.05

Note: All means and fractions are computed with the analysis population as observation units.

3. Empirical strategy

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 and present individuals in Norway aged 30-60 at this point in time. After that, new individuals are included 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 are included.

We model individual sickness absence propensity by means of a multivariate hazard rate model. We distinguish between “minor” (acute) and “major” (potentially chronic) diseases, based on initially recorded diagnosis/symptoms. The distinction is made on the basis of the aggregate duration distribution of absence spells *by diagnosis*, such that diagnoses 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, while the group of major diseases is dominated by musculoskeletal and mental diseases. The resultant length distributions of absence spells are described in Table 3. Around 70 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. 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. The reasons for this are first, that some diagnoses comprise a wide range of health conditions and, second, that our classification is based solely on initial diagnosis. An important point to note is also that the major disease diagnoses comprise most of the hard-to-verify diseases; hence, to the extent that sickness absence insurance involves moral hazard problems, we would expect it to show up in major, rather than minor disease absence spells.

Table 3: Sickness absence spells- minor and major diseases

	Major	Minor
Number of spells 2001-2005	2 664 651	1 052 163
Mean length (days)	46.12	8.94
Share of total absence days (%)	92,9	7,1
Length distribution (days):		
minimum	1	1
5 percentile	2	1
10 percentile	3	1
50 percentile	16	5
90 percentile	129	15
95 percentile	222	24
maximum	365	365
Share right-censored spells (%)	2.86	0.44

There are three alternative states that an individual can occupy in our model; presence (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:

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

where x_{it} is a vector comprising all (time varying as well as time invariant) observed factors assumed to affect an individual *i's* hazard rates at time *t* and (v_{12i}, v_{32i}) is a time-invariant unobserved characteristic.

Once absent, individuals are subject to the risk of recovery and, hence, becoming present. Let $\{T_2, T_3\}$ be the stochastic durations of absence in state 2 and 3, respectively. The two single risk hazard rates are then defined and specified as follows:

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

where d_{it} is a vector describing the duration of an ongoing absence spell and (v_{21i}, v_{31i}) are time-invariant unobserved characteristics. Note that the classification of the two types of absence spells (minor and major) is based on initial diagnosis/symptoms only; hence we do not model transitions between minor and major disease absence spells. Changes in diagnoses during a sickness spell are disregarded.

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. There are two reasons why we do this. First, we are not able to identify the initial diagnosis for these spells; hence we cannot categorize them correctly. Second, 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.

To derive the likelihood function, 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

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

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.

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. Provided that the model is non-parametrically identified, this can be done without imposing functional form restrictions on the heterogeneity distribution. However, given the computational challenges associated with non-parametric maximum likelihood estimation, there is a tradeoff involved regarding the flexibility by which observed and unobserved characteristics can be represented in the model. We therefore pursue two complementary strategies in this paper; one in which we represent observed characteristics of the environment in a very flexible way – with both firm and panel doctor fixed effects – but disregard unobserved heterogeneity at the individual level, and one in which we represent the environment in a more parsimonious manner (replacing firm and panel doctor fixed effects with a limited number of firm an panel doctor characteristics), but incorporate unobserved heterogeneity among employees in a completely nonparametric fashion. These two models also serve to illuminate different aspects of sickness absence determinants. The model with firm and panel doctor fixed effects is well suited for providing a comprehensive descriptive decomposition of absence behavior in terms of, e.g., individual, workplace, and panel doctor components. However, it does not contribute to explaining why absence behavior varies across firms and panel doctors. By replacing the firm and panel doctor fixed effects with observed firm and panel doctor characteristics we may to some extent reveal the underlying causal mechanisms and shed light on what causes the differences between firms and panel doctors.

Our data do not include information regarding the employees' occupation. Firm fixed effects will therefore not only capture the impacts of work environments, but also of the types of work that dominate within the firms and, hence, the occupational composition of their workforces. To some extent, we can think of our two model specifications as the "between firms" model (without fixed effects) and "within firms" model (with fixed effects). The interpretation of the effects of individual characteristics (such as age, gender, social background etc.) is slightly different in these two models. While the between firms model will capture the "gross" impact of individual characteristics (including the impact that goes via workplace selection/occupational choice), the within firms model will capture only the "net" effect (conditional on workplace).

In the model without firm and panel doctor fixed effects, the joint distribution of unobserved employee heterogeneity is approximated by means of a discrete distribution. Let Q be 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 $L_i\left(v_l\right)$ 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 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). 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.

4. Main results

This section presents selected results from the estimated models. The two models contain around 1,300 and 50,000 parameters, respectively; hence we cannot describe the results in any detail. For the fixed effects model, we were also unable to compute standard errors (since the resultant covariance matrix contains around 2.5 billion elements). We focus on the results from the model without fixed effects in this section, but refer to the fixed effects model in cases were the point estimates from the latter model deviates significantly from the former. The fixed effects model is used extensively in the next section for simulation purposes. Complete estimation results for both models are available from our web page www.frisch.uio.no/docs/absenteeism.html.

As discussed in the previous section, the model without fixed effects was estimated with a non-parametric maximum likelihood procedure. The unobserved heterogeneity distribution of the preferred model ended up containing 27 support points.

4.1 Demography

The impacts of age are illustrated in Figure 2. The probability of entering into a sickness absence spell declines sharply with age up to around 45 years. It thereafter stabilizes (minor diseases) or rises (major diseases). The probability of recovering from an illness declines monotonously with age. Given that the health condition does not systematically improve with age, the high entry rates into sickness absence for young people have (at least) three possible interpretations. The first is that young people have lower thresholds for claiming sick, e.g., because of weaker social norms against exploiting the sickness insurance system. The second is that youths generally occupy junior positions with more strenuous work and less flexible work-hours. And the third is that workers with poor health are gradually sorted out of the workforce.

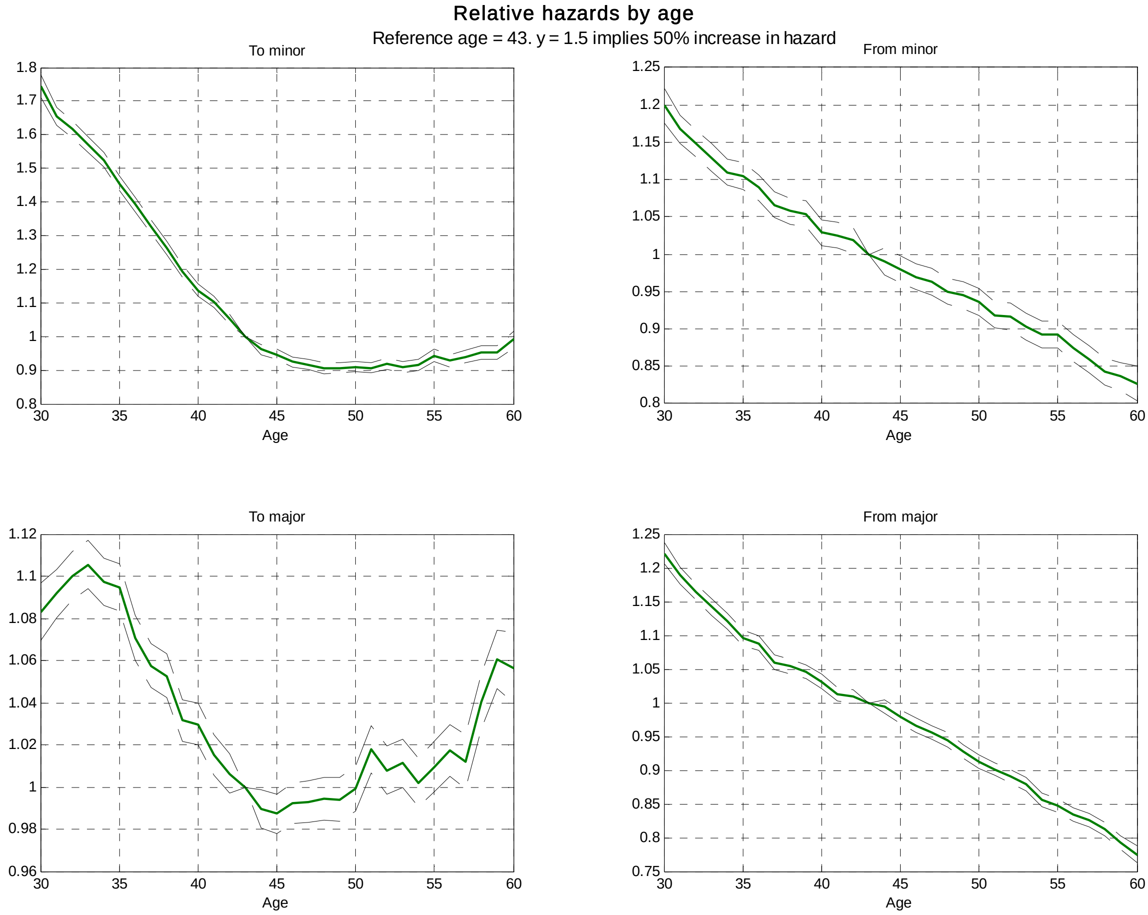


Figure 2. Estimated hazard rates by age (with 95 percent confidence intervals)

The age effects estimated in the fixed effects model were strikingly similar to those presented in Figure 2. From this we can infer that the age differences are hardly caused by sorting into firms by age. We can however not exclude that there may be sorting into *positions within* firms.

Women tend to have much higher entry rates and lower exit rates than men. In table 4 the impact of gender is displayed as the percentage shift in the hazard rates caused by being female rather than male. The non-parametric approach makes it possible to compare the gender differences, conditional on civil status and number of small children (below 12 years of age). Regardless of whether we compare married, separated/divorced or singles, with or without children, the gender differences are large. Women's entry rates to sickness absence exceed those of men by as much as 50-80 percent for minor diseases and 43-53 percent for major diseases. However, while these gender differentials are conditional on a number of individual characteristics, workplace characteristics, and work hours (see Table 1), there are no controls for occupation and job tasks. When we include firm fixed effects in our model the gender differential declines substantially. But female entry rates still exceed male entry rates by as much as 30-50 percent for minor diseases and 20-40 percent for major diseases, ceteris paribus.

The gender differentials in recovery rates depend strongly on whether we compare workers with or without small children. Without children, women tend to recover faster than men. With children, women recover substantially slower than men. For married men and women the differences in recovery rates are almost 8 percent.

Table 4: The impact of gender by marital status and number of children Females' additional hazard, in percent.

	Married			Separated/divorced		Never married	
	0 children	1 child	2 children	0 children	1 child	0 children	1 child
to minor	68	58.3	49.8	81.3	74.8	71.4	69.1
to major	43.3	48.5	45.7	46.2	53.6	50.6	43.7
from minor	1.4	-0.1	-0.9	3.7	2.0	2.9	2.3
from major	-0.1	-7.7	-7.9	6.4	1.1	-2.8	-2.9

There are large differences in absence patterns across different nationalities. Immigrants from EU or USA have entry rates to absence around 10 percent above those of natives. Immigrants from outside EU/USA have entry rates that are as much as 40-45 percent higher than natives. Recovery rates are fairly similar across nationalities, with the exception that immigrants from outside EU/USA have 16 percent higher recovery rates from major diseases than natives. The impacts of nationality become somewhat smaller when we include workplace fixed effects in the model. But non-western immigrants still have entry rates 30-40 percent above those of otherwise identical natives.

4.2 Family background

There is a strong social gradient in absenteeism. Even conditional on own education and income, 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 eight percent lower probability of obtaining a minor disease and almost 20 percent lower probability of obtaining a major disease than an otherwise identical person born in a family with both parents in the lowest education bracket (only compulsory education). Income differences between parents reinforce this social gradient, as the offspring's rate of entry into major diseases 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 rises by 30 percent.

4.3 Current family situation

Table 5 presents some key results regarding the impact of children and current family situation. It turns out that the effect of having children is somewhat different for men and women. For both married men and women, the presence of children increases the probability of entering a minor disease absence spell. The increase is as large as 10.2 percent for men and 3.8 percent for women, compared to having no children below 12 years. However, while the presence of small children seems to reduce the probability of entering a major spell and increase the probability of returning from any kind of absence spell for married men, the opposite is the case for married women. Note particularly that the recovery rate from major disease absence with children, compared to having no children (ceteris paribus) is -4.1 percent for females and +3.9 percent for men. This pattern is confirmed when we focus on separated/divorced.

Being separated or divorced entails a higher rate of entry into absence spells by as much as 17.5 percent for men and 25 percent for women, compared to being married. It also reduces the return probabilities by roughly 1.5 percent. Since separation/divorce is not randomly assigned, these statistical associations are of course not necessarily causal. We examine the causal impact of actually *becoming* separated/divorced in the next subsection.

When we compare single men and women with married men and women, both without children, there are some interesting differences. Single men are enters absence more seldom, and they return faster when entered, than married men. However, focusing on the major disease group, single women enter absence more often and return more seldom than married women.

Table 5. The impact of small children and marital status

Change in hazard rates associated with having one child below 12 years, percent.

	Married		Separated/Divorced		Never married	
	Men	Women	Men	Women	Men	Women
to minor	10.2	3.8	3.4	-0.3	18.0	16.5
to major	-1.6	2.0	1.0	6.1	12.2	7.0
from minor	1.3	-0.3	0.2	-1.4	0.1	-0.6
from major	3.9	-4.1	0.6	-4.5	-1.6	-1.7

Change in hazard rates associated with being either separated or single, compared to married, percent

	Without children below 12 years				1 child below 12 years			
	Separated		Never married		Separated		Never married	
	Men	Women	Men	Women	Men	Women	Men	Women
to minor	17.5	25.0	-0.7	-0.1	10.3	20.1	6.4	12.1
to major	17.4	24.5	-5.8	2.9	20.5	29.6	7.3	8.0
from minor	-1.3	-1.3	2.1	1.4	-2.3	-2.4	0.9	1.1
from major	-2.9	-1.4	5.6	-2.1	-6.0	-1.9	0.0	0.4

4.4 Family events

To illustrate the impacts of various family shocks, we focus on entry into the major diagnosis, as the shocks in question have little impact on the primarily infectious diseases defined as minor. Figure 3 illustrates the impacts of 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 absence entry rate prior to and after the event in question. These profiles are estimated non-parametrically, using dummy variables that capture the effect of particular time periods. There are consequently no restrictions regarding functional forms that drive any of the results.

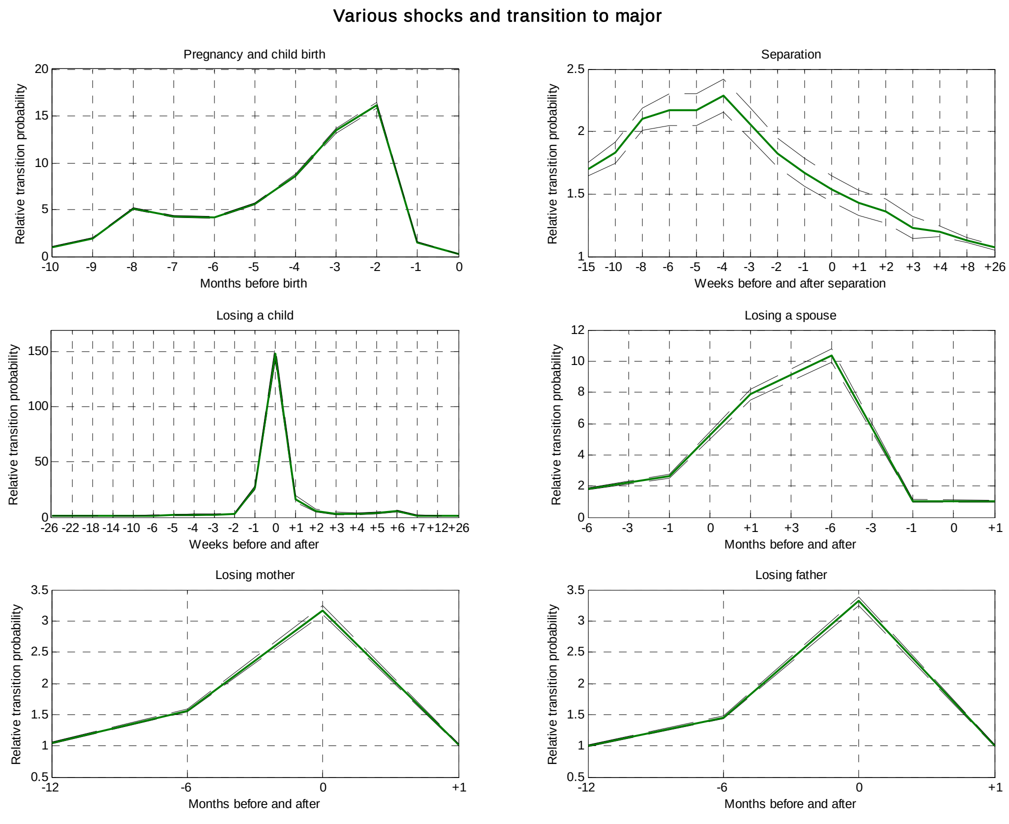


Figure 3. The impact of family shocks (with 95 percent confidence intervals).

Pregnancy has the almost immediate effect of five-doubling the rate of entry into major diseases; 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 2.5), and it declines rapidly after the couple has actually split. We recall from the previous subsection, however, that separated/divorced employees tend to maintain (indefinitely) a significantly higher entry rates into absence than married employees.

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.

4.5 Own human capital and job characteristics

Educational attainment appears to have a large impact on sickness absence; see table 6. In particular, we note that higher education sharply reduces the rate of entry into major diseases. The education differentials are reduced by roughly 25 percent when we include workplace fixed effects, suggesting that the education effect partly operates through the selection of occupation/workplaces.

Table 6
The impact of educational attainment relative to basic high school education. Percentage shift in hazard rates

	Elementary school only	High school finalized	High school additional	College/Univ. lower level	College/Univ. higher level	College/Univ. PhD level
to minor	6.6	-8.9	-18.1	-18.9	-34.9	-50.8
to major	11.6	-11.3	-23.7	-28.8	-48.3	-59.7
from minor	-3.1	0.7	2.3	2.6	5.8	1.8
from major	-2.8	1.3	0.0	-1.6	0.0	1.5

From our register data, we are unfortunately not able to separate completely between work hours and hourly wages. We observe overall earnings derived from a job and we observe whether the job is “small part-time”, “big part-time” or “full-time”. Within each of these categories, we divide the population into earnings quartiles. A key finding is that the entry rates into both minor and major diseases are highest among full-time workers in the lowest earnings-quartile. The absence entry rates generally rise with work-hours and decline with earnings (given full-time work). Hence, the overall absence rate tends to be lowest in the two tails of the overall earnings distribution, in the lowest tail because work-hours are very low and in the upper tail because the hourly wage rate (human capital) is so high.

4.6 Absence incentives

Given that the sickness insurance system pays a 100 percent replacement ratio to all absentees, there is no variation in the direct pecuniary incentives for avoiding absence. Yet, we suspect that the overall cost of being absent may be higher the less secure is the job. Hence, we expect individuals with temporary employment contracts to be less absent, ceteris paribus. Moreover, previous evidence indicates that local labor market conditions also affect the costs of absence since the threat of losing ones work is more frightening the poorer are the chances of getting a new job.

We use short tenure (less than one year) as a proxy for job insecurity. Needless to say, tenure is not randomly allocated; hence the estimated parameters are likely to reflect the causal effect of interest as well as the sorting into short tenure employment. As it turns out, short tenure is associated with a 6 percent *higher* entry rate into minor diseases and a two percent lower hazard rate into major diseases. The recovery rates are 3 percent higher for minor diseases and 5 percent higher for major diseases. We find it difficult to draw any substantive conclusions based on these findings.

Local labor market conditions are represented in the model by the monthly transition rate out of registered unemployment in the municipality. In the model without workplace and doctor fixed effects, the estimated impact of local labor market conditions will be largely driven by the cross-sectional variation between different geographical areas in Norway, while in the model with workplace and doctor fixed effects, they will be driven by idiosyncrasies in economic fluctuations within local labor markets. As it turns out, the two models provide very different answers to the issue of business cycle impacts. Without fixed effects, we estimate a negative impact of labor market tightness on entry into minor diseases, and a positive impact on entry into major diseases. With fixed effects, we find a positive impact on entry into minor diseases and no impact on entry into major diseases. Given that geographical variation in sickness absence behavior may arise from factors other than labor market conditions, we believe that the fixed effects model provides the most reliable estimates for the causal effect. However, given the lack of standard errors for this model, it is again difficult to draw firm conclusions.

4.7 The workplace

The composition of the workforce has a significant impact on each employee’s absence propensity. For example, we find that a balanced gender mix (within firm) implies lower entry rates into major diseases than both male-domination and female-domination. With respect to age, we find the opposite pattern. Entry rates are highest at workplaces with balanced age composition and lowest when the average age of colleagues is either very low or very high.

Workplace turbulence – as reflected by high turnover rates and high downsizing rates – typically causes small (1-2 percent) increases in subsequent absence entry rates.[4]

From October 2001, firms have had the opportunity to enter into the tripartite inclusive workplace agreement (IWA), implying that the threshold for GP certification of employees' sickness absence is raised from 4 to 9 days and that employers and employees commit on certain dialog procedures for each case of absence (with practical support from local "work-life-centers"). Given that firms' entry into the agreement is anything but random, the impacts of IWA should be evaluated on the basis of the fixed effects model. As it turns out, however, the two models yield very similar results. Entry into IWA implies a reduction in the employees' hazard to minor disease absence by around 8-9 percent and a corresponding reduction in the recovery rate by around 4 percent. It has no impact at all on entry into or exit from major disease absence. Our interpretation of these findings is that the only impact of IWA affiliation is that it reduces the need for obtaining certification for absence caused by minor (infectious) diseases.

4.8 Sorting of current workforce

A high local employment rate may be taken as an indicator that even 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 indeed that a higher local employment rate implies high entry rates into both types of sickness absence. On the other hand, it also implies high recovery rates, suggesting that the marginal members of the workforce have more frequent – but also less serious –

[4] Note that we only focus on the *subsequent* impacts here, i.e., the turnover and downsizing variables are lagged to avoid endogeneity problems.

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 – somewhat surprisingly – 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 at all confirm the idea that high employment necessarily entails a high rate of sickness absence.

4.9 The panel doctor

It turns out to be difficult to pinpoint particular panel doctor characteristics that are able to explain the variation in hazard rates. The gender and age of the panel doctor seem to be irrelevant. Whether the GP has a specialist education and/or shares office with other colleagues is also irrelevant. However, the doctors' economic incentives seem to play a role. In particular, panel doctors with a significantly shorter patient list than desired tend to certify 4-6 percent more absence spells than panel doctors with full lists. This may indicate that doctors become more lenient in response to patient shortage. The GP's certification practices also seem to be affected by the competitive pressures in the municipality. The larger is the average patient shortage in the municipality, the higher is the employees' entry rates into sickness absence. The difference between doctors in the highest and lowest municipality shortage quartile is around 7 percent for entry into both minor and major disease absence spells. This again indicates that patient shortage puts the GP's in a weaker "bargaining position" towards the patient.

4.10 Time and duration dependence

Figure 4 displays the calendar time pattern in the four hazard rates (generated from quarterly time dummy variables). There is a strong seasonal pattern in the entry rates into sickness absence, particularly for minor diagnoses; see Figure 4. There is no obvious time-trend in absence behavior during the period covered in this paper, perhaps with the exception that recovery rates from major diseases have deteriorated somewhat. There also seems to have been a shift in entry rates around the summer of 2004, coinciding with the reform of the sickness absence certification regulations discussed above. Both entry rates reached historically low levels in the first quarter after the reform (2004.3), with the decline in entry rates to major diseases being most conspicuous.

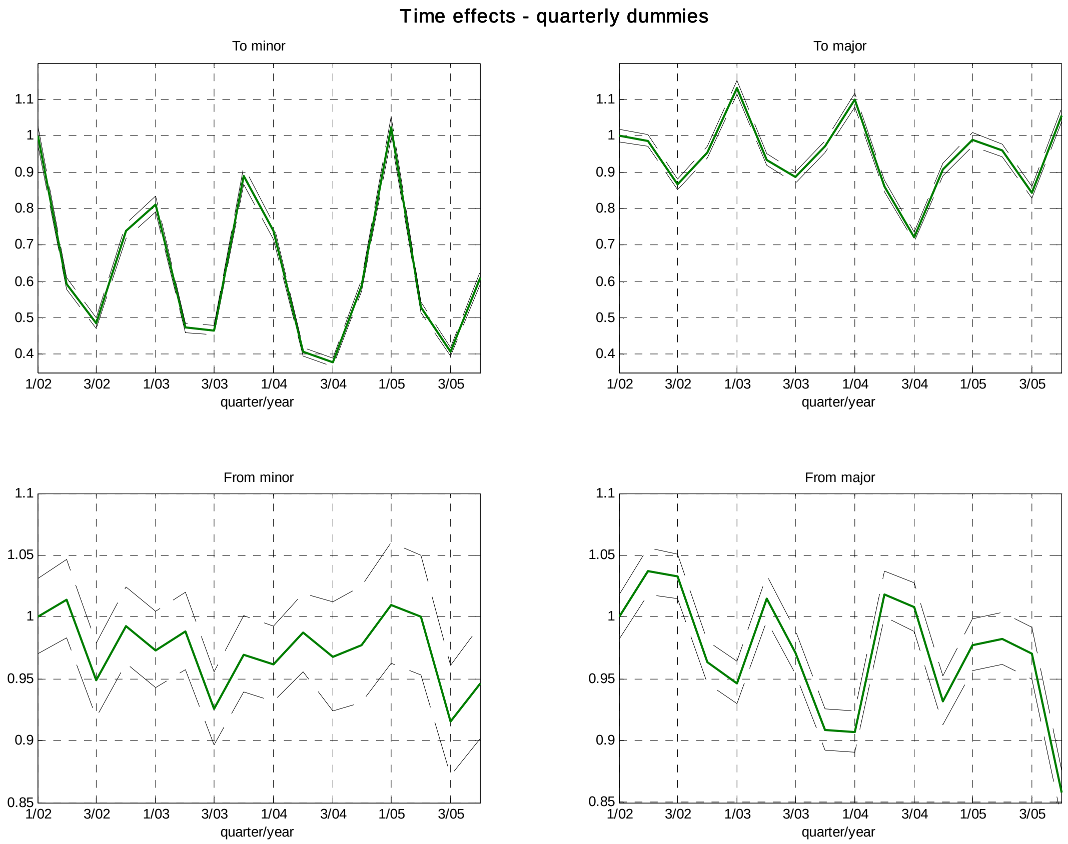


Figure 4. Calendar time effects. Hazard rates are rescaled relative to 1st quarter 2002. Dashed lines are 95% confidence intervals.

The calendar time developments in recovery rates in Figure 4 are evaluated at the beginning 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 observed average recovery frequency during the first absence week. The graphs illustrate that there is strong negative duration dependence in recovery hazards. For example, the probability of recovering from a major disease declines from around 26 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.

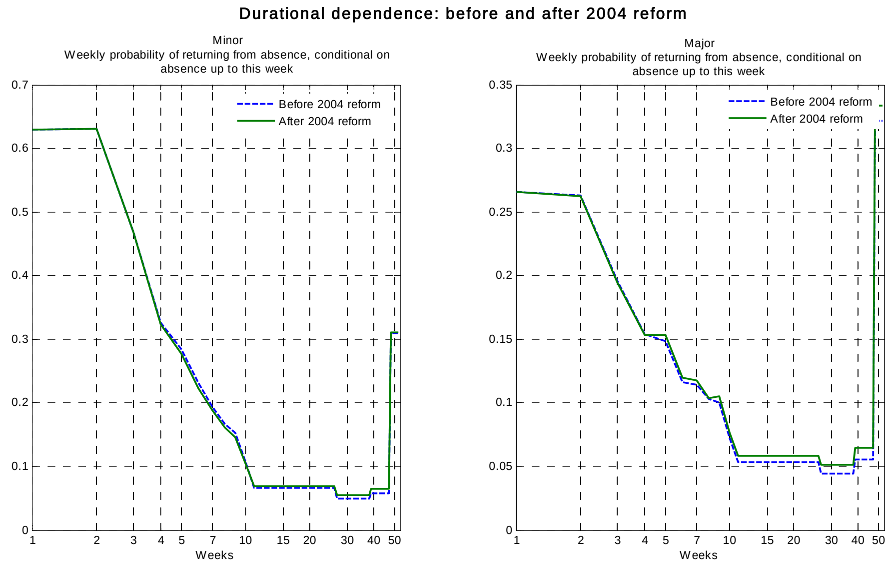


Figure 5. Duration dependence in recovery hazards. Before and after the reform in 2004.

As referred to above, Figure 5 also illustrates the impact of the reform in 2004. This reform made work presence 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 return probabilities from the major disease group 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.

Our model also included a time-indicator for important sports events. Interestingly, these events raised males’ entry rates to major diseases by around 8 percent, while leaving the females’ entry rates unchanged.

5. Decomposition of sickness absence propensity

In this section we use the estimated models to simulate absence behavior among Norwegian employees, and we then use the simulation results to decompose the variation in sickness absence propensity into its main sources.

5.1 Longitudinal variation

TBW

5.2 Cross-sectional variation

In order to examine the sources behind the cross-sectional variation in sickness absence propensity, we build on the fixed effects model. This model makes it possible to quantify the impacts of firms and GPs relative to individual determinants. It is not obvious, however, how sickness absence propensity should be defined. The cross-sectional variation in sickness absence at a given *point in time* is clearly not of particular interest, since all individuals are then either absent or not absent, and little information is revealed regarding the underlying absence propensities. This reflects the more general phenomenon that the shorter the time-window we look at, the less information is revealed regarding each agent’s underlying sickness absence propensity. In a short time-period, the cross sectional variation in actual absenteeism is also amplified by the random nature of health shocks. As the time-period becomes longer the random component becomes less important, and each employee will approach an average absence rate corresponding to his/hers intrinsic absence propensity (regardless of the initial state). We therefore use the estimated hazard rate profiles to compute the *steady state absence level* for each employee in the dataset at a particular point in time. This is done by using the limiting distribution of the markovian transition matrix (Taylor and Karlin, 1998, p. 207).[5] There are at least two ways to think about the somewhat abstract concept of a person's steady state absence rate. First, it 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. Second, one can think of it simply as a convenient summary representation of underlying hazard rates.

Figure 6 (the two lower panels) illustrates the extent to which the variance of observed absence rates (in the data) declines with the length of the observation period. At the bottom of the two panels, we also show how the variance declines further when we substitute our steady state absence rates for the observed absence rates. Note, however, that these steady state rates are calculated from the fixed effects model which does not account for the presence of unobserved heterogeneity. Steady state rates have been calculated under two alternative assumptions regarding transitory shocks (such as divorce and death in close family); i) that they last indefinitely, and ii) that they did not occur at all.

The variance in predicted steady state absence rates illustrates the variation that our model is able to explain. In steady state nothing is random and all differences

[5] 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.

are due to differences in covariates between people. If we compare steady state variance with the variance in the five year data sample, we get an idea of how much of the total variation the model is able to explain. If we assume that none of the variation in the five year sample is due to randomness, the share of the variance explained by the model gives a goodness of fit measure, a quasi R-squared. The model is able to explain 5 percent of the between-employee variation in the minor disease absence and 13 percent of the variation in major disease absence (without family shocks included). This constitutes a lower bound on the model's explanatory power as some variation in the five year sample probably is due to randomness.

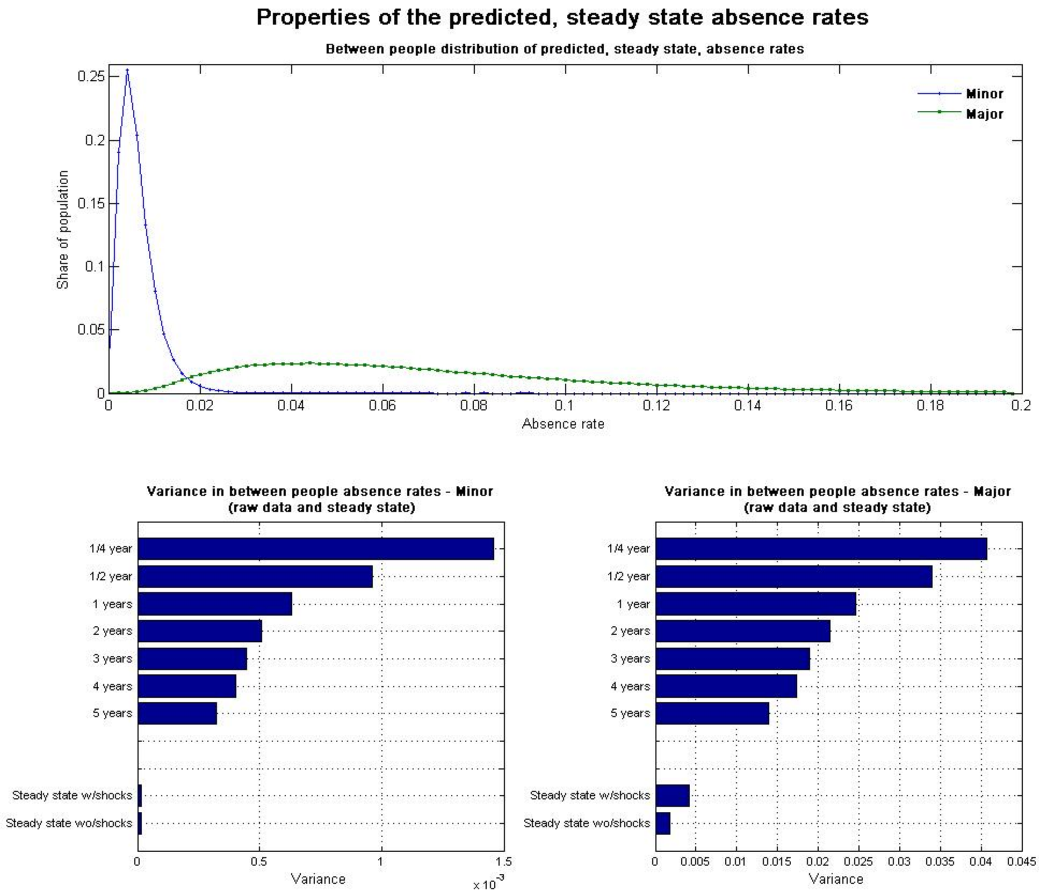


Figure 6. Predicted, steady state absence rates. The upper panel illustrates the distribution of absence, minor and major, in the workforce. The lower panels illustrates how the variance in absence rates (between people) drops as the time frame is extended.

The upper panel of figure 6 shows the distributions of predicted steady state absence rates in the minor and major categories. It is striking how different they are. Virtually all employees have a steady state minor disease absence rate between 0 and 2 percent. The steady state rates of major disease absence, on the other hand, are spread relatively evenly out between 1 and 10 percent, and a significant probability mass remains at absence levels as high as 16 percent.

We now turn to the issue of decomposing the variation absence rates that our model does account for into its various sources, and to disentangle sorting mechanisms from causality. Figure 7 illustrates one way of tracing out the roles of firms and panel doctors. The figure plots the distribution of mean absence rates for firms with more than 50 employees (upper panel) and panel doctors (lower panel) in the five-year data period (cubes), in the model-based simulations (stars) and in a modified model-based simulation where all variables except firm and panel doctor dummies are replaced by their respective means (circles). The differences between the latter two distributions reflect the impacts of employee sorting. A first point to note is that the model's steady state absence rates quite nicely reproduce the distribution of mean absence rates across firms, while they fail to accurately reproduce the distribution of mean absence rates across panel doctors. A second point to note is that employee *sorting* seems to be an important factor behind the variation in absence rates across firms, but that it is of more limited importance for the variation in absence rates between panel doctors. Finally, it is clearly the case that there is large variation in absence rates across both firms and panel doctors even after controlling for the composition of employees. This indicates that the government's focus on policies aimed at affecting firm and GP behavior may have some merit.

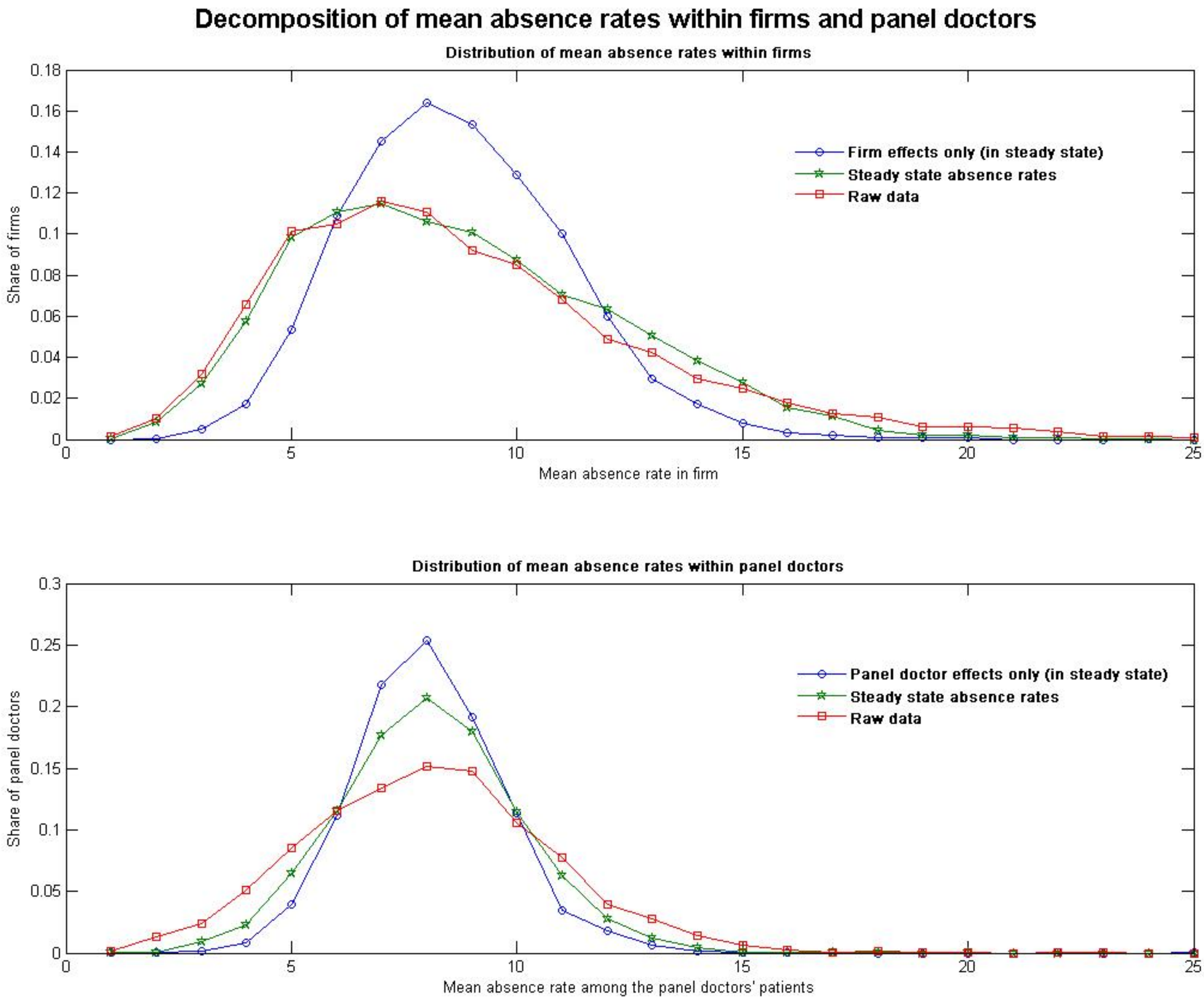


Figure 7. Variation between firms and panel doctors. Pure randomness ,sorting and causal effects.

One way of summarizing the empirical importance of different groups of explanatory variables for the overall variation in absence behavior is to examine their impact on the overall cross sectional variance of latent absence rates. Let a_{ki} be individual *i's* latent (steady state) absence rate corresponding to state *k* (*k=2,3*). The variance of predicted absence rates $(\operatorname{var} \hat{a}_{ki})$ can be decomposed into its various components in the following way: $\operatorname{var} \hat{a}_{ki} = E\left[\operatorname{var} \hat{a}_{ki} \mid C\right] + \operatorname{var}\left[E\left[\hat{a}_{ki} \mid C\right]\right]$, where C is some conditioning statement, e.g., firm or panel doctor. The fraction of total variance accounted for by differences in the conditioning statement is thus $\frac{\operatorname{var}\left[E\left[\hat{a}_{ki} \mid C\right]\right]}{\operatorname{var}\hat{a}_{ki}}$. This will reflect a combination of sorting on the conditioning variables and their possible causal effects. For example, the fraction of absence variance accounted for by gender partly reflects that women and men end up in different types of jobs and partly that gender has a causal impact, ceteris paribus. Our estimated model gives us the opportunity to disentangle the variance contributions into their sorting and causal components simply by computing steady state rates with and without various blocks of causal effects removed. Intuitively, the variation between, e.g., firms after removal of the firm effects must be due to selection of workers into firms, and not the firms them selves. The results from a number of such decomposition exercises are presented in Table 7.

Table 7. Sorting versus causality.
Fraction of variance in predicted latent absence rates accounted for by different groups of variables, with and without their causal effects included, in percent

	Full model		Model with causal effects of the variables in question removed		"Causality share"		Causal impact Variance between groups caused by this variable	
	I Minor	II Major	III Minor	IV Major	[I-III]/I Minor	[II-IV]/II Major	I-III Minor	II-IV Major
Gender, children and civil status	22.6	25.5	1.81	5.68	0.92	0.78	20.8	19.8
Age	3.8	5.8	1.3	0.9	0.65	0.85	2.5	4.9
Education	16.4	38.9	3.8	13.5	0.77	0.65	12.6	25.4
Workplace	53.0	54.2	24.1	28.9	0.55	0.47	29.0	25.4
Panel doctor	32.4	22.5	7.6	12.8	0.77	0.43	24.8	9.7
County	4.3	4.1	4.0	3.2	0.07	0.22	0.3	0.9
Parents' health	4.0	8.4	1.3	1.9	0.67	0.77	2.7	6.5
Parents' education and income	5.7	11.9	4.1	8.3	0.29	0.30	1.66	3.62
Wage and work hours	8.7	19.8	3.5	3.6	0.60	0.82	5.2	16.2

The results indicate, for example, that roughly half of the overall variation in predicted steady state absence rates is accounted for by firms. However, half of this variation is caused by the fact that different firms have different workers. Hence, around a quarter of the overall absence rate variance is causally explained by the firms (and the type of work that is carried out within them). Among panel doctors, roughly 25-30 percent of the variation is between panel doctors. Around 25-30 percent of the absence variation is accounted for by panel doctors. The causality share is 77% for minor spells, and 43% for major spells. This indicates that the doctor has a major impact on whether or not workers are sick-listed for minor illnesses, but limited influence on the large bulk of spells in the major disease category.

6. Conclusion

TBW

7. References

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)

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. (1984b) 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.

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.

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

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