DISCUSSION PAPER SERIES

IZA DP No. 5343

The Case for Presenteeism

Simen Markussen
Arnstein Mykletun
Knut Røed

November 2010

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Case for Presenteeism

Simen Markussen
Ragnar Frisch Centre for Economic Research

Arnstein Mykletun
Norwegian Institute of Public Health

Knut Røed
Ragnar Frisch Centre for Economic Research and IZA

Discussion Paper No. 5343
November 2010

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. 5343
November 2010

ABSTRACT

The Case for Presenteeism*

Can activation requirements control moral hazard problems in public sickness absence insurance and accelerate recovery? Based on empirical analysis of Norwegian data, we show that it can. Activation requirements not only bring down benefit claims, they also reduce the likelihood that long-term sickness absence leads to inactivity. Our findings show that absentees who are issued graded (partial) absence certificates by their physician have shorter absences and higher subsequent employment rates than they would have had on regular sick leave. We conclude that the activation strategies that in recent years have permeated European and US welfare policy may fruitfully be carried over to sick leave insurance.

JEL Classification: C26, I18, I38, J48

Keywords: sick leave, disability, activation, workfare

Corresponding author:

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

* This research has received support from the Norwegian Research Council (grants no. 185201 and 187924). Thanks to Bernt Bratsberg and Oddbjørn Raaum for comments and discussions. Thanks also to seminar participants at the University of Utrecht, the Norwegian school of economics and business administration (NHH), the University of Bergen, and the University of Oslo.

1. Introduction

Public insurance against income losses during sickness absence from work involves a well-known moral hazard problem: Heavily insured workers tend to be absent too often. And firms that can pass their insurance costs on to the public purse exert too little effort to prevent it. This is obviously costly for those who pay the insurance premium (typically the taxpayers). In addition, it potentially involves large costs for the absent workers in the form of slower recovery and ensuing earnings losses. Recent medical research indicates that for the illnesses responsible for the vast majority of sick leave days in advanced economies – such as musculoskeletal pain and common mental disorders – regular activity through work helps promote recovery and rehabilitation; see, e.g., Waddell (2004) and Waddell and Burton (2006). And empirical labor market research shows that frequent and/or longer term absence spells significantly reduces subsequent employment and earnings prospects (Hansen, 2000; Ichino and Moretti, 2009; Markussen, 2010), and thus potentially also raises the probability that a worker becomes inactive and dependent on social insurance payments on a lasting basis. Disability benefit recipiency is rising inexorably in many industrialized countries, typically with long-term sick leave as the major entryway. And at present, 6% of the OECD working-age population receives a quasi-permanent disability benefit (OECD, 2009, p. 3).

To the extent that the level of absenteeism is considered to exceed its socially optimal level, possible remedies would be to cut the level of sickness benefits and/or to hold employers accountable for a larger share of insurance costs (e.g. through some form of experience rating). But these options may be blocked, either by competing political prior-

ities or by legally protected contractual obligations. In this paper we show that there is yet another way towards both lower absenteeism and less labor market exclusion, namely to impose *activity requirements* in the sickness insurance system. Most sick leave days are caused by non-communicable diseases for which it is far from obvious that 100 % absence from work is the appropriate treatment. Insurance systems should therefore be designed to break the "mechanical" link between the presence of pain and symptoms and absence. In particular, we argue that long-term sickness episodes rarely justify complete inactivity. Sickness normally reduces an individual's work-capacity, but it rarely eliminates it. This calls for *graded (partial)* sickness insurance arrangements, i.e. insurance that covers the loss arising from reduced productivity or work-hours due to illness, while requiring the worker to exploit his/her remaining work-capacity. Graded sickness insurance thus promotes *presenteeism*, i.e., that workers are present at their workplace even when they are sick, but of course only when the illness is non-infectious and otherwise compatible with work. This idea represents a significant extension of the *activation strategy* that has dominated both European and US welfare and unemployment insurance policies in recent years. With some variations, the use of graded sickness insurance has recently been strongly promoted in the Nordic countries (Kausto *et al.*, 2008), and it has also been introduced in the UK in the form of a so-called "fit note".[1]

We provide empirical evidence from Norway suggesting that for workers on long-term sick leave (more than 8 weeks) activity requirements not only reduce absenteeism

[1] The UK *fit note* (Statement of Fitness for Work) was proposed in 2008 and implemented in April 2010. In the fit note, physicians are requested to certify whether a sick worker is unfit or (potentially) fit for work. In the latter case doctors may recommend reduced hours or duties, and provide recommendation to employers on how they can help the worker back to ordinary work.

and social security benefit claims, but also significantly improves their subsequent employment prospects. The evidence is based on recent attempts by policy makers in Norway to make graded absence the *default* option during sickness. Physicians writing medical certificates to employees have been encouraged to prescribe graded rather than complete absence as much as possible. And for absence spells exceeding 8 weeks, they have been instructed to certify complete absence only in exceptional cases. So far, however, there has been no systematic monitoring of physicians' practices and no sanctions against physicians prescribing (too much) non-graded absence. In the present paper, we exploit the fact that there has been a huge variation in physicians' degree of compliance with the activation-strategy, generating a significant source of random-assignment-like (from the employee's point of view) variation in the probability of being subject to activity requirements during spells of sickness. We apply an instrumental variables technique, where we use physicians' observed tendency to grade *other patients'* sick leaves as the instrument. We examine the consequences of incorporating different groups of control variables, including physician characteristics such as workload, market situation, and overall propensity to issue absence certificates (leniency), as well as neighborhood fixed effects to remove any geographical sorting. To assess the model's reliability, we also estimate it on outcomes for which we should not expect any causal effects, i.e., *past* absence and employment for current absentees and future absence and employment for workers with no sick leaves at all (but who consult the same physicians).

The key finding of our paper is that the use of graded rather than non-graded sickness absence certificates reduces the length and volume of long-term absence spells and significantly improves the likelihood that the absentees are employed in subsequent

years. The effects are large, both from an economic and a clinical perspective. Our most conservative instrumental variables estimates indicate that switching from a non-graded to a graded absence certificate before the 12th week of absence reduces the length of the absence spell by as much as 70-85 fulltime-equivalent days and raises employment propensity two years after by 12-14 percentage points for both men and women. Our findings thus indicate that the introduction of activation requirements in sickness insurance schemes may be a promising strategy towards reducing sickness absence insurance costs and combating labor market exclusion.

2. Existing literature

Our paper relates loosely to a large literature on the impacts of workplace-based return-to-work (RTW) interventions (see, e.g., Franche *et al.*, 2005, for a recent review), and more directly to a small literature on the impacts of graded sick leave. The former literature demonstrates that workplace interventions tend to improve sick-listed workers' chances for returning to work. However, these interventions typically entail treatment far beyond the "prescription" of return to (some) work, e.g., in the form of physiotherapy, cognitive-behavioral interventions, organizational changes, etc. They are also typically targeted at workers with particular diagnoses, such as musculoskeletal disorders. Hence it is difficult to draw general conclusions regarding the isolated impacts of graded versus non-graded absence certification. There is some corroborating evidence from clinical trials showing that the recovery prospects of back pain patients and individuals with light mental disorders may be enhanced by continuation of "normal activities" even without additional treatments. Malmivaara *et al.* (1995), for example, conducted a controlled trial

among Finnish employees with acute nonspecific low back pain. The patients were randomly assigned to one of three treatments: bed rest for two days, back-mobilizing exercises, or continuation of ordinary activities "as tolerated". It turned out that the latter ordinary-activity-group had significantly faster recovery than the other two. A Norwegian randomized controlled trail of long-term absentees with lower back pain also found that advice to stay active was associated with better prognosis for return to work during three years follow-up than "treatment as usual" in primary health care (Hagen et al, 2003). A recent literature review for the UK Department of Work and Pension concluded that work for sick and disabled people is therapeutic and leads to better health outcomes (Waddel and Burton, 2006).

We are aware of two previous attempts to identify the causal effects of graded sick leave, both limited to examining the impacts on sick leave duration itself. The first is documented in a series of working papers based on Swedish register data; see, e.g., Andrén and Andrén (2008; 2009). Since the authors use non-experimental data, they face an obvious endogeneity problem in that graded and non-graded absence certificates are not randomly assigned. This problem is handled by means of an instrumental variables approach. The authors apply the workers' occupation as instrument, arguing that while occupation heavily influences the potential for working reduced hours, it is unlikely to influence the recovery prospects and, hence, the overall length of the absence spell directly.[2] Based on this identifying assumption, they find that graded sick leave reduces the speed of recovery during the first 3-4 months of absence, whereas it speeds it up after-

[2] This identifying assumption is arguably questionable. Occupation may be correlated to sick leave duration for several reasons other than its impact on the grading propensity, both related to the nature of the occupations (e.g., physical demands) and to the sorting of employees into occupations.

wards. The second attempt to estimate the effect of graded absence is based on Danish register and survey data (Høgelund *et al.*, 2010). Here, the identification problem arising from non-random assignment is handled by means of a mixed proportional hazard rate model and the timing-of-events approach (Abbring and Van den Berg, 2003), thereby avoiding the need for instruments. Identification then relies on the validity of the proportional hazards and no-anticipation assumptions (the latter requires that workers do not anticipate a forthcoming transition from fulltime to graded absence). The findings suggest that graded instead of non-graded absence raises the weekly probability of returning to regular hours by as much as 50 percent.

In addition to these two causal analyses, there is also a number of studies on attitudes towards graded sick leave and of self-reported patient experiences. Most of these have been released in the form of non-peer-reviewed reports and working papers. According to a recent review (Kausto *et al.*, 2008) the results mainly indicate positive attitudes towards graded sick leave, among employees, employers, physicians, and social security administrators. One (published) study examining the subjective views of sick-listed workers in Sweden (Sieurin *et al.*, 2009), for example, report that 92 per cent of the workers on graded absence, and 63 percent of the workers on non-graded absence consider graded absence to be (potentially) "good for me".

The present paper adds to the existing literature both by introducing a new and powerful source of identification – i.e., the variation in practice styles across primary care physicians – and by examining the impacts on a much wider range of outcomes than what has been previously offered.

3. Institutional setting

Norwegian workers are entitled to a 100 percent replacement ratio from the first day of sick leave and up to one year. The first 16 days are paid for by the employer, the remaining days are paid for by the social security administration. The only limitation is that a general practitioner (GP) must certify absence spells exceeding 3 days (8 days in some firms). During periods of sickness absence, Norwegian workers enjoy a special protection against dismissals, implying that they cannot be dismissed on grounds that are related to their sickness.[3] The moral hazard problems are fairly obvious in this case. Workers have incentives to be absent more than necessary. Firms have incentives to make efforts to prevent short-term absence (since they cover the full costs during the first 16 days), but not necessarily to reduce long-term absence. Indeed, it is typically more beneficial for the firm that a long-term absentee continues to be absent than that he/she returns to work with a high risk of again becoming sick. Moreover, given the level of employment protection in Norway, firms with excessive labor may sometimes find it convenient that they can pass the costs of temporary redundant labor on to the social security administration. And after one year of sick leave, continued absence becomes a legitimate cause for dismissal.

Norway also has a high level of absenteeism. On a typical working day, around 7 percent of all workers are absent due to sickness. Long-term sick leave entails a high risk of labor market exit and continued social security dependency. At the point of sickness

[3] The burden of proof lies with the firm. In practice, this implies that absent workers can only be laid off as part of a mass displacement. After the one year absence period, the firm is allowed to lay off the worker if the sickness implies that he/she is no longer able to perform his/her duties.

benefit exhaustion after one year, around 65 percent of the claimants move on to temporary disability benefits (medical or vocational rehabilitation), typically with a replacement ratio around 66 percent; and 3 years after exhaustion, 30 percent have become permanently disabled.[4] More than 20 percent of the working age population in Norway is now dependent on a health-related social security transfer (Bratsberg *et al.*, 2010). And in the National Budget for 2011, public insurance payments for sickness absence and disability are projected to account for 5.1 percent of GDP.

Despite the high level of absenteeism, *all* political parties in Norway, as well as the associations of employers and employees, agree that the existing replacement ratio should be maintained, and that no additional costs should be passed on to firms with absent workers. Instead, policy makers have chosen to focus on "softer" measures, such as more intensive use of graded absence certificates, public information campaigns, and support for improvements in workplace environments. Given that almost 90 percent of all absence days in Norway are certified by a physician, certification practices have received considerable attention. In 2001, a tripartite "inclusive workplace agreement" (IWA) was made between the state and the associations of employers and employees, in which a target of 20 percent reduction in absenteeism within four years was set. The agreed strategy included the encouragement of substituting graded for non-graded sick leaves, particularly for long-term absence spells. Graded sick leave implies, for example, that if a worker's work-capacity is considered to be temporarily reduced by 50 % due to an illness, he/she is obliged to work at 50 % capacity and entitled to the normal wage for this part. If the

[4] Source: Own calculations based on all sick leave spells starting in 2002.

spell exceeds 16 days, the remaining 50 % are paid for by the social insurance system.[5] During the sickness period, the employer is obliged to facilitate modified work within reasonable limits, while the employee is – if necessary – obliged to accept changes in regular duties/tasks. Graded sick leave normally implies part-time absence (reduced work hours), but in principle it can be implemented in the form of less productive work (with unchanged hours) also. Given the costs associated with implementing workplace adaptations, it is not intended for very short sickness absence spells; hence short-term absence certificates are typically not graded. If the physician expects an absence spell to be long-lasting, it can nevertheless be graded from the start. And for non-graded absence certificates covering sick leaves beyond 8 weeks, the physician is obliged to explain to the social security administration why grading cannot be used. Typical explanations are that it is difficult to implement the required workplace adaptations or that the sick leave spell is expected to end very soon anyway.

There are three ways in which grading is intended to affect remaining absence duration and subsequent outcomes (Mykletun *et al.*, 2010): First, it is aimed at pushing unmotivated workers back to work, even when health problems prevent them for working at full capacity, thereby containing moral hazard problems among employees. Second, it is expected to coerce employers to make appropriate efforts to facilitate adapted work, thereby containing moral hazard problems among employers. And third, it is intended to improve the employees' health and to speed up their recovery.

[5] Note that the maximum duration of a sick leave spell is 1 year regardless of its grade.

In response to relatively modest use of graded sickness absence certificates during 2002 and 2003, a reform in the absence-certification-regulations was implemented in July 2004, which, *inter alia*, explicitly instructed physicians to use graded absence certificates for all long-term absence spells (exceeding 8 weeks) unless the spell is expected to end shortly or work-related activity is directly harmful to the health of the employee or his/her colleagues. Figure 1 shows that while the total level of physician-certified absence trended upwards in 2002 and 2003, it declined sharply around the time of the 2004 reform.[6] Since the reform also contained a number of other elements, it is difficult to use it directly to identify the causal effects of grading. The impacts of the reform are evaluated by Markussen (2010), who concludes that the reform did cause a drop in absenteeism by more than 20 percent, and that the increased use of graded absence certification was one of the probable causes for the reform's apparent success. Yet, the actual rise in the use of graded absence certificates was moderate; the fraction of spells that was graded at least at some time during the sick leave rose from around 15 to 20 percent for all sick leave spells and from 25 to 40 percent for spells exceeding 8 weeks. Hence, the aim of making graded absence the "default" – and non-graded absence the exception – was not achieved.

Norway has, since 2001, practiced a family (panel) doctor system, whereby each citizen is assigned a single physician who receives a capitation fee from the social securi-

[6] Norwegian attempts at promoting the use of graded absence certificates were inspired by similar earlier efforts in Sweden. In Sweden, the fraction of absence certificates that are graded rose from around 25 percent in the 1990's to well over 30 percent after the turn of the century. It peaked in 2007 at a rate of 37.2 percent. And interestingly, also in Sweden, the rise in the grading fraction turns out to be mirrored in a decline in the overall absence rate; see www.forsakringskassan.se/press/statistik_och_analys.

ty authorities. Sickness absence certificates can in principle be issued by any authorized physician, but in cases of long-term sickness, it will normally be issued by the family doctor (except when the patient is hospitalized or subject to intensive specialist treatment). Norwegian workers are free to choose their family doctor insofar as the physician in question has vacant patient slots. As we return to below, this implies that we face a potential endogeneity problem when we use family doctor characteristics to instrument absence certificates.

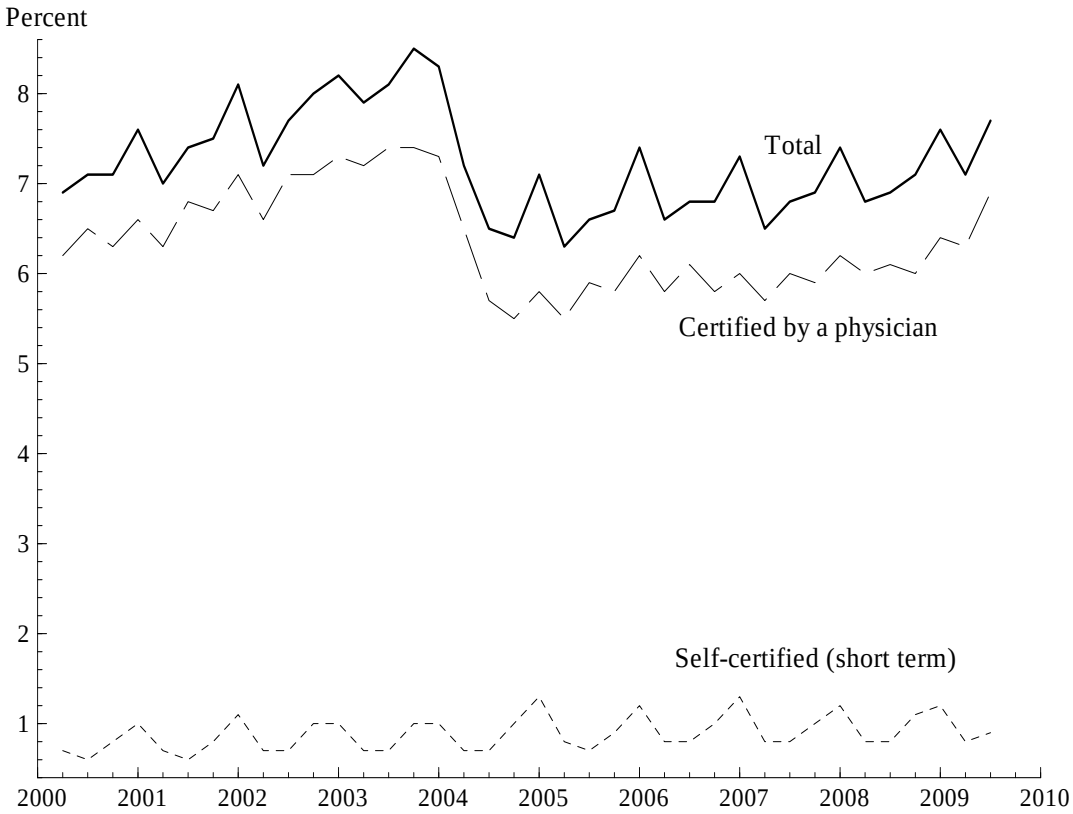

Figure 1. Sickness absence in Norway 2000.2-2009.3 (percent of agreed work-hours) Source: Statistics Norway.

4. Data

The data we use are collected from Norwegian administrative registers and include encrypted information about all citizens and their primary care physicians from 2001 and onwards. The data include detailed longitudinal information on employment and social security spells, and annual information on earnings. They also include longitudinal information on all certified absence spells from 2001 through 2005, including starting and stopping dates, diagnosis and grade (graded or not graded).[7] By merging different administrative registers (employer-employee, education, demography, social security, income/taxes) we are able to obtain ample information about each patient.

In the present analysis, we exploit data on all long-term absence spells (more than 8 weeks) in Norway from 2001 through 2005 handled by the family doctor (390 580 spells). A spell is recorded as graded if a partial absence certificate is issued before the 12th week.[8] Table 1 summarizes our patient data. Graded absence certificates were issued in around 29 % of the cases, and it was used much more frequently for women than for men. Patients with graded certificates were slightly older, had slightly higher education, and had higher earnings prior to the absence spell than patients with non-graded certifi-

[7] Note that we only use grading information as a dichotomous variable (graded or not graded). We do not exploit information on the actual grade. The reason is that there is not enough variation across physicians in grading percentages to identify the effects of actual grade on the basis of an instrumental variables approach. The most commonly used grade is 50 %, which is used in around 60 % of the graded absence certificates.

[8] Since we are going to evaluate the impact of grading on absence duration, this induces a small reverse causation problem. The probability of obtaining a graded certificate is higher if the spell lasts 12 weeks than if it lasts 8 weeks. We have nevertheless chosen to include grading decisions up to the 12th week for the reason that many transitions to graded absence occur in this period. To the extent that reverse causation is a problem it will bias the estimated impact of grading on absence duration *upwards*, which is in the opposite direction of the effects we are going to find in this paper.

cates. Table 1 also presents the key patient *outcomes* that we intend to focus on in the empirical analysis:

i. The total number of days from the start to the stop of the absence spell (including holidays and days off).
ii. The number of lost fulltime equivalent working days during the absence spell (i.e., the total number of days adjusted for expected days off (two per week), regular work-hours, and absence grade).
iii. The number of additional fulltime equivalent days on social security during the 24 months following the start of the long-term absence spell (caused by new absence spells, medical and vocational rehabilitation, unemployment benefits, social assistance, or permanent disability).
iv. Employment in the second year after the start of the absence spell (e.g., if the spell started in 2001, employment is evaluated in 2003).[9]

The descriptive statistics in Table 1 show that patients with graded absence certificates on average had much more favorable outcomes than patients with non-graded absence certificates. They had shorter absence durations, lower degree of social security dependency afterwards, and higher subsequent employment rates. Differences in outcomes between patients with graded and non-graded absence certificates represent a combination of sorting and causality. Physicians obviously take the patients' prospects

[9] A person is interpreted as employed in a year if earnings from regular or self employment exceeded approximately 120,000 NOK ($ 20,000) in 2005-value.

Table 1. Descriptive statistics

	Men		Women	
	Grad. absence certificate	Non-graded absence certif-icate	Grad. absence certificate	Non-graded absence certif-icate
# Observations (long-term spells)	35 984 (22.1%)	126 668 (77.9%)	74 271 (33.6%)	146 738 (66.4%)
I. Patient characteristics (year *t*)				
Age	44.3	42.3	44.5	43.5
Years of schooling	12.8	12.1	13.2	12.5
Initial annual earnings (NOK)	375 657	349 705	298 840	265 024
II. Outcomes				
Mean absence duration (days)	148.4	201.5	151.1	214.4
Mean number of fulltime-equivalent sick leave days	67.4	132.5	60.1	117.3
Mean number of additional fulltime equivalent days with social security dependency next two years	72.5	106.8	70.3	93.4
Employed year *t+2*	91.6	78.7	92.1	76.8

Note: Year *t* is the calendar year in which the long-term absence spell starts. Initial annual earnings are measured in year *t*. These earnings are unaffected by the absence spell since sickness benefits are included in our earnings measure and the replacement ratio is 100% during the first year

into account when making their grading decisions; hence the differences between physicians' grading propensities may mirror differences in their patient composition. In particular, the patterns in Table 1 may reflect that physicians issue graded certificates disproportionally to employees with favorable prospects. A person who is too sick to work at all will obtain a non-graded absence certificate and at the same time probably have a low likelihood of a quick recovery. On the other hand, a graded absence certificate normally entails non-trivial costs/efforts for the employer and the employee who need to agree on the required workplace adaptations, e.g., in the form of changes in work-hours as and job contents. For this reason, an absence spell is typically not graded if it is expected to be

very short-lived.[10] This mechanism induces a negative correlation between grading propensity and future prospects. Which of these sorting mechanisms that dominates for the spells covered by our analysis (all spells exceeding 8 weeks) is an empirical question.

We intend to disentangle causality from sorting by exploiting the variation in grading-propensities across physicians. There are 3 868 family doctors included in our analysis. On average, each of them issued 100 long-term absence certificates during our data window (29 graded, 71 non-graded). The variation in actual use of graded absence certificates was substantial. This is illustrated in Figure 2, panel A, where we have divided the physicians into 20 equally sized groups based on their grading frequency, for male and female patients, respectively. While some physicians almost never use graded absence certificates, others use it in 80-90 % of the cases. This does of course *not* prove that physicians have different grading-propensities; the observed variation in grading frequencies could in principle be fully explained by patient sorting.[11] We return to this issue in the next section. Looking at the data, however, there is a remarkable correlation between the physicians' observed grading frequencies and their patients outcomes; see Panels B-E in Figure 2. The graphs indicate strong and monotone relationships between physicians' grading propensity and their patients' subsequent outcomes. The higher their physician's grading propensity, the better the patients perform along all outcome dimen-

[10] While graded absence certificates tend to be associated with favorable outcomes for the long-term sick leave spells analyzed in this paper, they are associated with negative outcomes among all spells. The average duration of all non-graded spells in our data was 35 days. In comparison, the average duration of graded spells was 85 days.

[11] Existing evidence suggests, however, that variation in clinical practice between physicians is an important determinant of expenditure for primary physician services in Norway. Grytten and Sørensen (2003) show that, depending on the type of diagnosis, physician-specific effects explained 47-66 % of the variation in expenditure for laboratory tests and 59-66 % of the variation in expenditure for consultations lasting over 20 minutes.

sions. Thus, if there is a causal component in this pattern, the potential for cutting absenteeism and prevent labor market exclusion by means of graded absence certificates may be substantial. The remainder of this paper seeks to contain the sorting problem and, hence, to isolate and estimate the causal effects of grading decisions on the four outcomes illustrated in Figure 2.

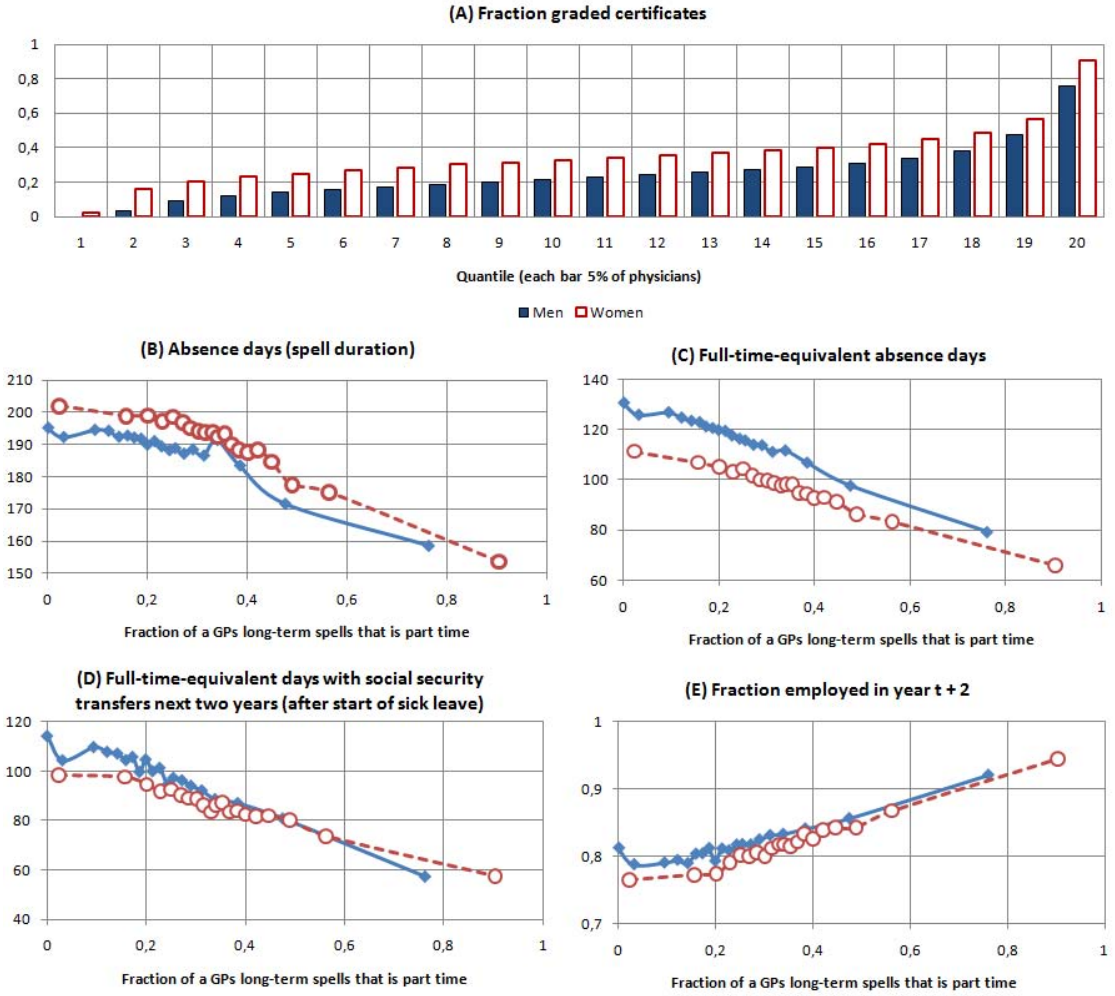

Figure 2. Grading fraction of long-term spells by quantile in the physician grading distribution (panel A) and observed patient outcomes by the physicians' overall grading frequency (panels B-E).

5. Empirical Analysis

Consider a situation where a physician faces the choice between certifying a *graded or a non-graded* absence spell for a particular worker, where the former decision implies that the worker continues working to the extent deemed tolerable.[12] The physician's decision may affect the length and the ultimate outcome of the worker's absence spell, and hence also his/her future employment prospects. The purpose of our analysis is to quantify these effects empirically within the population of workers for whom the physician's discretion actually influences the choice of treatment (the type of absence certificate).

Let y_i be one of the outcomes for worker *i*, such as absence duration or subsequent employment. The regression equations of interest can then be written

$$y_i = x_i'\beta^S + \alpha^S P_i + \varepsilon_i, \quad S = M(ale), F(emale), \tag{1}$$

where x_i is a vector of conditioning observed variables (including a constant term), P_i is an indicator variable taking the value *1* if the absence certificate prescribes graded (partial) absence (and *0* if it prescribes non-graded absence), and ε_i is an unobserved residual. All the parameters are allowed to vary by gender, both because it is plausible that the impacts of grading are different for men and women and (as shown in Table 1) because the use of graded certificates is much more common for female than for male patients. The parameters of interest are (α^M, α^F). Now if we had experimental data – with controlled randomization of P_i – we could estimate Equation (1) directly with an appropriate

[12] The physician can obviously also decide not to certify any absence spell at all, and in some cases, this may be the most realistic alternative to a graded absence certificate. We abstract from this complication at the present stage, but return to it below.

statistical model (or simply compare the outcomes for patients with full and graded absence certificates). Since we do not have experimental data, we must take into account that P_i is not likely to be independent of ε_i. A physician obviously takes the employee's recovery prospects into account when making decisions about the absence certificate, and it is unlikely that we can find observed control variables (x_i) that fully captures the information available to the physician. To eliminate this source of endogeneity bias, we pursue an instrumental variables strategy. Our instrument is going to be a variable that we can think of as the physician's *practice style*. The idea behind this instrument is to exploit the variation in the usage of graded absence certificates that we observe between physicians, after having controlled for patient characteristics. While this can be shown to constitute a powerful instrument – in the sense that it strongly affects an employee's likelihood of obtaining a graded sickness absence certificate – it is more difficult to ascertain that it is completely independent of the residual (ε_i) in the outcome equations. There are two mechanisms by which dependency may arise. The first is unobserved patient-physician sorting, which may arise both due to the geographic location of physicians and patients and through endogenous choice of family doctor. The second is that physicians who are different with respect to their usage of graded absence certificates may be different along other dimensions as well, e.g., in their overall propensity to issue absence certificates ("leniency") and in the quality of their medical advices and treatments. The next sub-sections explain how we have dealt with these problems and how we have ascertained empirically – by means of robustness exercises – that the problems have actually been appropriately dealt with. We start out presenting our instrumental variables strategy.

5.1 The instrumental variables model

Let $J = 1,...,N_J$ be the set of family doctors in Norway. Assume that a patient's probability of obtaining a graded rather than a non-graded absence certificate can be written as a function of all factors that potentially affect the outcome of the absence spell x_i and of his/her physician's grading propensity G_j, i.e.,

$$\Pr(P_i = 1 \mid x_i, D_i) = x_i'\theta^S + D_i'G, \quad S = M, F, \tag{2}$$

where D_i is a vector of dummy variables with the *j'th* element equal to 1 if patient *i* has physician *j* as his/her family (panel) doctor, and other elements equal to zero, and G is the corresponding vector of physician grading propensities. If G had been observed, it would clearly have been the obvious choice of instrument for P_i in Equation (1). Since it is not, we instead estimate it. A natural way of obtaining estimates for G is to estimate Equation (2) with the realized certificate outcome P_i as the dependent variable, applying a separate dummy variable for each physician represented in the dataset, and then take the dummy-coefficients as estimates for the respective elements of G (or, equivalently, use a vector of physician dummy variables as instruments directly). However, this strategy suffers from a non-ignorable "reflection problem" (Manski, 1993) caused by the fact that each worker's grading outcome (P_i) contributes to the estimation of his/her physician's grading propensity; hence the resultant instrument is not orthogonal to ε_i. We deal with this problem by estimating the grading-propensities on the basis of a completely different patient-population than the one for which we use it as instrument. More specifically, we apply the grading-propensities computed from the female patient-population to instru-

ment males' grading outcomes and vice versa. Estimating Equation (2) by means of OLS for men and women, we obtain two estimators for each physician's grading propensity $(\hat{G}^{-F}, \hat{G}^{-M})$, where the superscripts (*-F,-M*) indicates that the indexes have been estimated without females and males represented in the dataset, respectively. Let $\hat{g}_i^{-S} = D_i' \hat{G}^{-S}$; i.e., a scalar variable attributing to patient *i* the estimated grading propensity of his/her physician (where the estimate is based on the opposite sex population). The first step equations in our instrumental variables approach thus become:

$$P_i = x_i' \delta^S + \lambda^S \hat{g}_i^{-S} + \xi_i, \quad S = M, F. \tag{3}$$

The reduced form outcome equation takes the form

$$y_i = x_i' \beta^S + \varphi^S \hat{g}_i^{-S} + \zeta_i, \quad S = M, F, \tag{4}$$

and the instrumental variables estimator for α^S in Equation (1) is

$$\alpha_{IV}^S = \frac{\hat{\varphi}^S}{\hat{\lambda}^S}, \quad S = M, F. \tag{5}$$

The reliability of our instrumental variables approach depends on whether the conditional independence assumption holds; i.e., whether $\varepsilon_i \perp \hat{g}_i^{-S} \mid x_i$. In the present context this is not obvious, since patient-physician sorting may be similar for men and women and since the grading indicator may be correlated to other relevant physician characteristics. To ascertain conditional independence, we apply a comprehensive list of control variables (x_i), and we examine the robustness of our results with respect to the inclusion of various variable sets. The potential variable sets include *patient and job characteristics, time indicators, neighborhood indicators, and physician characteristics*. A complete de-

scription of these variable sets is provided in the Appendix. Here, we only briefly explain their contents and the role they are designed to play in our analysis.

Patient and job characteristics are included to control for observed patient-physician sorting. The list of observed characteristics is extensive, and includes age, nationality, education, industry, work-hours, and present earnings, as well as past employment, earnings, and absence (during the preceding three years). To avoid unjustified functional form restrictions, we have chosen to represent most of the variables in the form of extensive dummy-sets. Neighborhood indicators are included to contain any remaining unobserved residence-based patient-physician sorting. There are 12 921 "neighborhoods" in Norway represented in our dataset. These neighborhoods are typically small and homogeneous; on average there are only 30 absence spells in each neighborhood during our five-year observation window. Controlling for them implies that we exploit the within neighborhood variation in explanatory variables and outcomes only. Time-indicators are included to ascertain that correlated time trends in patient-physician sorting and outcomes (e.g., related to the 2004 reform) are not allowed to affect our results. To avoid unnecessary restrictions, we include a separate dummy variable for each of the 56 possible starting months in our data window. Finally, we include physician characteristics to control for differences in practice styles beyond their grading propensity.[13] The vector of physician characteristics include observed variables, such as gender, age, specialization, number of patients, and the rate of vacant patient slots. But, importantly, it also con-

[13] Note that we generally use the same vector of control variables when we compute physicians grading propensities (Equation (2)) and when we evaluate their impacts (Equations (3) and (4)). The only exception is the model where we include physician characteristics, since these characteristics obviously cannot be included together with physician dummy variables in Equation (2).

tains additional practice style indicators particularly designed to reflect their *leniency* – in terms of their willingness to issue any kind of absence certificate (to capture any systematic differences in their patients' health). These indicators are based on auxiliary regressions and additional datasets covering all Norwegian employees. They are computed in a similar fashion as the grading propensities, but based on different datasets and outcomes. In essence, they are designed to reflect the physicians' propensity to issue short-term and long-term absence certificates to their employed customers during a year; see the Appendix for details.[14]

5.2 Estimation results

The main estimation results are presented in Table 2, based on alternative conditioning sets. We also report results based on Ordinary Least Squares (OLS), with all available control variables included, for comparison. A first point to note is that the physicians' estimated grading propensity (based on the opposite-sex-population) has a significant impact on the likelihood of obtaining a graded absence certificate; conf. the first-stage coefficients. Hence, our instrument is strong, and it is approximately equally strong for men and women. The second stage estimates indicate that the grading decision has large and lasting impacts on patient outcomes. By grading long-term absence certificates, physicians contribute to shorter absence durations, less subsequent social security dependen-

[14] We also include a proxy indicator for the quality of the physicians' medical advices. We do this by exploiting yet another cut of our register data consisting of all elderly retirees in Norway. The indicator is computed by regressing patient mortality on physician dummy variables, controlling for all observed patient characteristics. Although it can be argued that the physicians' influence on their patients' mortality is too small for this approach to capture the genuine quality differences between physicians, we believe that this is the best we can do with our data to obtain an indicator that at least may be (weakly) correlated with treatment quality.

cy, and higher employment propensities later on. Regardless of which of the models we rely on, the effects are large from an economic viewpoint. Grading the absence certificate at week 8-12 reduces the expected number of (fulltime-equivalent) work-days lost during the spell by as much as 70-85 for both men and women. Since the average numbers of lost days for non-graded spells were 132 for men and 117 for women (conf. Table 1), this implies that grading is estimated to cut the overall absence volume by half. In addition it reduces subsequent social security claims and raises employment. For men, the number of saved social security days after the sick leave spell (but within a two-year period after the start of the spell) is estimated to be of similar magnitude as the number of days saved within the spell, whereas for women it is somewhat lower. Taken together, these estimates imply that if the overall grading propensity were doubled (from 29 to 58 % in our sample), expected social security payments during the first two years after the start of a long-term absence spell would be cut by 19 %. The potentially most important impact of grading, however, is that it significantly raises the probability of remaining in employment afterwards. Even according to the most conservative of our IV-estimates, issuing a graded rather than a 100 % absence certificate raises the employment by around 12-14 percentage points for both men and women.

Comparing the different IV-estimators, we note that the results are only moderately affected by changes in the sets of conditioning variables. Exploiting the within-neighborhood variation in grading decisions and outcomes only (columns C and G) rather than the national variation (columns B and C) does little to change the results for women whereas it reduces the estimated employment effects for men. Adding physician characteristics in the regression – including indicators for physician leniency – has only marginal

effects on the estimates (columns D and H). The latter suggests that it is indeed the physicians' grading strategies that drive our results; and not differences in absence certification practices *per se*. A particular point to note is that our findings debunk the idea that high grading-propensity implies that the threshold for issuing absence certificates becomes lower. To the contrary, our first-stage estimates indicate that there is a significant negative correlation between a physicians' leniency and the probability of obtaining a graded absence certificate (not shown in the table). This may suggest that a physician's focus on grading actually makes sick leave a *less attractive option* both for employees who don't really need it and for employers who with some effort are able to prevent it.

Table 2. Estimation Results (robust standard errors in parentheses, clustered on physicians)

	Men				Women			
	A	B	C	D	E	F	G	H
I. First stage:								
Effect of excluded instrument	-	0.309 (0.012)	0.240 (0.012)	0.225 (0.013)	-	0.277 (0.016)	0.192 (0.011)	0.174 (0.011)
F excluded instrument	-	361.5	257.5	225.5	-	291.8	176.8	153.2
II. Second stage: Effect of grading								
Absence duration	-56.1 (0.74)	-39.6 (12.5)	-48.9 (15.6)	-45.4 (16.0)	-64.9 (0.59)	-17.7 (12.7)	-59.3 (16.3)	-54.1 (17.7)
# Fulltime-equiv. lost working days in abs. spell	-65.0 (0.41)	-83.6 (7.9)	-84.6 (9.9)	-82.6 (10.2)	-66.0 (0.31)	-67.4 (6.3)	-83.3 (8.5)	-81.0 (9.3)
# Fulltime-equiv. soc. sec. days next two years	-23.8 (0.73)	-104.6 (12.8)	-91.8 (16.0)	-90.2 (17.2)	-22.0 (0.50)	-61.8 (10.8)	-45.4 (14.5)	-35.9 (15.5)
Employed in year *t+2*	0.092 (0.002)	0.218 (0.038)	0.144 (0.047)	0.143 (0.049)	0.096 (0.002)	0.125 (0.034)	0.136 (0.045)	0.123 (0.050)
OLS/2SLS	OLS	2SLS	2SLS	2SLS	OLS	2SLS	2SLS	2SLS
Patient and time controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Neighborhood fixed effects	Yes	No	Yes	Yes	Yes	No	Yes	Yes
Physician characteristics	Yes	No	No	Yes	Yes	No	No	Yes
# Observations	162652	162652	162652	162652	221009	221009	221009	221009

Note: Year *t* is the calendar year in which the long-term absence spell starts.

Although the point estimates vary somewhat from model to model, the key lessons coming out the analysis are the same regardless of model specification.[15] We thus conclude that activation requirements during episodes of long-term sickness are highly effective with respect to containing social security expenditures and preventing processes of premature labor market exits. The large effects must be understood in terms of the strong incentives that *both* employers and employees otherwise have to pass sick leave costs on to the social security system; see Section 3.

Comparing the 2SLS with the OLS results, we find that there is a tendency for 2SLS to yield *larger* effects for all outcomes except absolute duration. This may suggest that the unobserved sorting into graded absence certificates is generally negative – in the sense that patients with poor outcome prospects have higher grading likelihood than patients with more favorable outcome prospects. A likely explanation for this is that grading is generally not used if the health problems are expected to end shortly; conf. the discussion in Section 4. However, IV-regression not only differs from OLS in that it removes sorting bias; it also bases the effect estimates on a more limited group of patients, namely those whose grading outcome is manipulated by the instrument, see Angrist and Krueger (2001, p.77). Our 2SLS estimates can thus be given a local average treatment effect (LATE) interpretation. It is conceivable that the complier-group – those whose grading outcomes are influenced by the physician – are workers of relatively good health, but

[15] This conclusion holds for a number of alternative model specifications not reported in the paper. We have also estimated the model on a reduced dataset consisting of physicians responsible for more than 10 long-term certificates only (and their patients), and the results were very similar to those reported here. Finally, we have experimented with different ways of splitting the population for the purpose of estimating the grading indicator on a different population than the one used to evaluate its impact. When we split the population randomly into two equally large groups (instead of by gender), we obtain estimates that are between those reported for men and women here.

with an uncooperative employer and/or limited own motivation, implying that the grading decision serves the dual purposes of containing moral hazard problems and promoting healthy work. By contrast, for motivated "always-takers" – i.e., patients who actively ask for a graded rather than a non-graded absence certificate – it is likely that only the latter effect is relevant.

5.3 Robustness and reliability

The results from the alternative model specifications in Table 2 indicate that our main findings are fairly robust, even though point estimates vary somewhat with the different conditioning sets. One problem that we have not addressed so far, however, is that of endogenous physician selection: Maybe unmotivated workers who are offered a graded absence certificate (only) respond by choosing a new physician, rather than by going back to work? This could generate a pattern of patient-physician sorting such that the motivated workers end up with the physicians with high grading propensity. One way to examine the empirical relevance of this argument is to look at the pattern of physician changes in the period around the issuing of graded and non-graded absence certificates. Table 3 shows that there was indeed a higher tendency for employees with non-graded absence certificates to have changed physician just before or during the sick leave spell than for employees with graded certificates. There was also a slight tendency for patients to move towards physicians with low grading propensity in the period around the issuing of a long-term absence certificate. The numbers and differences are small, however, in relation to the significant impacts reported in the previous subsection.

Table 3. Physician changes before (1 month) and during sick-leave spells. By patients grading status and physicians' grading propensities (standard errors in parentheses)

	Men				Women			
	Actual grading				Actual grading			
	Graded		Non-graded		Graded		Non-graded	
Percent changing before spell	0.94		1.18		0.89		1.09	
	(0.05)		(0.07)		(0.03)		(0.04)	
Percent changing during spell	1.88		2.88		1.29		2.00	
	(0.03)		(0.05)		(0.03)		(0.04)	
	Physician's grading propensity By quartile				Physician's grading propensity By quartile			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Percent changing before spell	1.45	1.03	1.02	1.13	1.49	0.91	0.81	1.10
	(0.07)	(0.04)	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)	(0.05)
Percent changing during spell	3.41	2.61	2.36	2.43	2.72	1.58	1.33	1.82
	(0.10)	(0.07)	(0.07)	(0.09)	(0.08)	(0.05)	(0.05)	(0.06)

Table 4. Estimation results based on patients without change in family doctor (robust standard errors in parentheses, clustered on physicians)

	Men		Women	
	Full model (Column. D, Table 2)	Non-switchers	Full model (Column H, Table 2)	Non-switchers
I. First stage:				
Effect of excluded instrument	0.225	0.217	0.174	0.168
	(0.013)	(0.013)	(0.011)	(0.011)
F excluded instrument	225.5	202.3	153.2	147.2
II. Second stage: Effect of grading				
# Absence duration	-45.4	-39.7	-54.1	-56.6
	(16.0)	(16.7)	(17.7)	(18.0)
# Fulltime-equiv. lost working days	-82.6	-79.6	-81.0	-79.1
	(10.2)	(10.8)	(9.3)	(9.5)
# Fulltime-equiv. soc. sec. days next two years	-90.2	-83.7	-35.9	-30.7
	(17.2)	(17.8)	(15.5)	(16.2)
Employed in year *t+2*	0.143	0.129	0.123	0.104
	(0.049)	(0.052)	(0.050)	(0.051)
OLS/2SLS	2SLS	2SLS	2SLS	2SLS
Patient and time controls	Yes	Yes	Yes	Yes
Neighborhood fixed effects	Yes	Yes	Yes	Yes
Physician practice style indicators	Yes	Yes	Yes	Yes
# Observations	162652	156937	221009	215327

Note: Year *t* is the calendar year in which the long-term absence spell starts.

To investigate the possible bias arising from physician-switchers, we re-estimate the model (with all control variables included) on the subset of non-switchers. The results are reported in Table 4. They show that estimated coefficients change very little. Hence, "physician-shopping" just before and during the absence spell does not seem to be an important part of the story.

To further evaluate the reliability of our model, we estimate it on a set of *past outcomes* for which the grading decisions under study could not possibly have had any causal effect. Since we have used patient history during the three years prior to the start of the spell in our control variables set, we use observed outcomes four and five years prior to the start of the absence spell to test the model. We focus on incidences of long-term absenteeism and employment in this exercise.[16] Note, however, that data limitations prevent us from using exactly the same outcomes for absenteeism (we do not have the same detailed data prior to our observation window). Even though a grading decision taken in year *t* obviously have not affected patient outcomes in years *t-4* and *t-5*, we cannot completely rule out that impacts of physicians' consistent activation strategies show up in past outcomes as well (resulting from past grading decisions). If this is the case, however, it implies that our IV-estimator exaggerates the effects of single grading decisions, whereas the reduced form effects (the second stage estimators multiplied by the first stage estimates) still correctly represents the impact of physicians' grading practices.

Estimation results for past outcomes are presented in Table 5. The point estimates indicate that a grading decision in year *t* implies slightly lower sick leave probability for

[16] The outcome capturing other types of benefits (rehabilitation and disability benefits) is not relevant in this context, since they are typically dependent on sickness insurance benefits being exhausted.

men and slightly higher employment probabilities for women in years *t-4* and *t-5*. Although none of the estimates are statistically significant at conventional levels, the pattern of estimates may indicate that the reduced form estimates do capture some grading effects beyond the single grading decision under study. The population examined in this study is clearly characterized by serious health problems (given their absence spells of minimum 8 weeks), and it seems probable that some of them have consulted their physician at many occasions prior to the absence spell examined here.

Table 5. Estimated second stage effects on past outcomes (robust standard errors in parentheses, clustered on physicians)

	Men	Women
Outcomes four years before (*t-4*)		
Incidence of long-term sick leave	-0.073	-0.005
	(0.059)	(0.065)
Employment	0.038	0.067
	(0.031)	(0.037)
Outcomes five years before *(t-5)*		
Incidence of long-term sick leave	-0.079	-0.037
	(0.058)	(0.067)
Employment	0.023	0.073
	(0.036)	(0.046)
Patient and time controls	Yes	Yes
Neighborhood fixed effects	Yes	Yes
Physician practice style indicators	Yes	Yes
# Observations	161748 (*t-4*)	220258 (*t-4*)
	161040 (*t-5*)	219576 (*t-5*)

Note: First stages are the same as in columns D and H, Table 2. Year *t* is the calendar year in which the long-term absence spell starts.

As a final check for the existence of systematic unobserved patient-physician sorting – related to "physician-shopping" and/or unobserved confounders – we exploit the population of workers that did not have any physician-certified sickness spells at all during the period from *t-2* to *t+1*, not even of short term. The idea behind this exercise is that the *t+2*-outcomes for these workers cannot possibly have been much affected by their

physicians' grading propensities; hence they can serve as a robustness test for remaining unobserved patient-physician sorting. We use *t*=2003 as the basis for this exercise, since this is the only year for which we can impose both the [*t-2*,*t+1*] no-absence restriction and observe the *t+2* outcome.[17] The results are presented in Table 6. The 2SLS estimator is of course meaningless in this case, since we look at a population with no sick leave spells; hence we report the reduced form estimates. To make these comparable to our main results reported in Table 2, the latter can be transformed to reduced form coefficients by multiplying the second stage coefficients with their first stage counterparts. For example, the estimated reduced form effect of the physicians' grading propensity on employment in *t+2* for the male population of long-term absentees is $0.14 \times 0.2 = 0.028$ (see Table 2). In comparison, the effect on the no-absence population is 0.001. As it turns out, all the estimated reduced form coefficients for the non-absent population are close to zero and statistically insignificant, just as we would expect if our control variables are sufficient for taking care of the systematic patient-physician sorting on future employment and absence propensities.

[17] Repeating this exercise for multiple years would in any case largely involve the same population.

Table 6. Estimated reduced form effects of physicians' grading propensities (robust standard errors in parentheses, clustered on physicians)

	Men	Women
Outcomes two years after (*t+2*)		
Employment	0.0014	0.0049
	(0.0059)	(0.0035)
Incidence of sick leave	0.0010	0.0009
	(0.0117)	(0.0067)
Incidence of long-term sick leave (>8 weeks)	-0.0063	-0.0008
	(0.0042)	(0.0027)
Patient and time controls	Yes	Yes
Neighborhood fixed effects	Yes	Yes
Physician practice style indicators	Yes	Yes
# Observations	265279	137132

6. Concluding remarks

Our estimation results indicate that requiring workers on long-term sick leave to be active – and work to the extent deemed tolerable by their physicians – is a highly effective tool to reduce social security dependency and raise subsequent employment. The effects are significant, both from an economic and a statistical viewpoint. Our most conservative estimates imply that substituting a graded for a non-graded absence certificate cuts the number of lost (fulltime-equivalent) working days during the sick leave by half, and raises the employment propensity two years later by 12-14 percentage points. Although the point estimates vary somewhat across different instrumental variables model specifications (with different conditioning sets), the overall evidence presented in this paper overwhelmingly supports the hypothesis that activation requirements do have the intended favorable effects on social security dependency and employment. We also find no evidence whatsoever that the use of graded absence certificates reduces the threshold for

claiming sickness benefits. To the contrary, physicians who frequently use graded absence certificates also seem to issue fewer long-term absence certificates in the first place.

Balancing the objectives of appropriate social insurance and sufficient work-incentives is a difficult task. For unemployment insurance and social assistance programs, policy makers in many countries have to an increasing extent resorted to various *activation strategies*, essentially requiring benefit claimants to participate in temporary employment or education programs. The key idea behind this strategy – with potential appeal to the political right as well as to the left – is that by pairing insurance with activity requirements it becomes possible to partly escape the unpleasant tradeoff between equality and work incentives; i.e., it facilitates a reduction of the moral hazard problem, given the level of insurance, or, alternatively, to improve the insurance coverage, given the level of moral hazard. The results presented in the present paper suggest that the same strategy may successfully be pursued for publicly provided insurance against income losses due to sickness absence. When workers with health problems are required to exploit remaining work capacity in order to be eligible for insurance payments, it probably becomes less attractive to report sick when it is not strictly necessary. It also becomes more difficult for employers to "get rid of" workers with health problems. Moreover, recent empirical evidence indicates that work is actually a *healthy* activity for workers with the illnesses and symptoms responsible for the vast majority of absence days in industrialized countries (musculoskeletal diseases, back pain, and light mental disorders). Although it could be argued that patients have every opportunity to take the expected adverse long-term consequences of inactivity into account when determining their own absence behavior, it is not difficult to imagine that some of them fail to do so, either because of insuf-

ficient information or limited self-control. In any case, the existence of generous sick leave insurance and the absence of experience rating imply that the decisions of employers and employees will be distorted to some extent.

What we have shown in this paper is that more intensive use of graded absence certificates has the potential of massively reducing the volume of social security insurance payments caused by sickness, and also of strengthening the employees' subsequent labor market attachment. There are obviously also some potential costs involved that we have not looked into in this paper. In particular, graded absence may involve workplace adaptations that impose costs on employers and/or co-workers. Physicians' assessments of sick workers remaining work capacity are also likely to be imprecise, implying that the insurer will not typically compensate the true productivity loss caused by health problems. Given the large gains involved, we nevertheless conclude that incorporating activation requirements into sickness insurance systems stands out as a promising avenue for future social security reform.

Appendix

Overview of control variables

The list of controls used in the regression analysis include 295 variables, most of them dummy variables. In addition, in some of the models, the data are centered on 12 921 neighborhood dummy variables. A list of all control variables used in this paper is provided in Table A1

Table A1. List of control variables used in the regression analysis

Patient characteristics
Age: One dummy variable for each yearly age (40 dummy variables)

Education: One dummy for each type-level education (70 dummy variables)

Industry: One dummy for each industry, two-digit NACE (61 dummy variables)

Nationality: One dummy for each region of origin (8 dummy variables)

Earnings level: One scalar variable with earnings in year *t* (inflated to 2005-value)

Work-hours: One indicator for each of three working time arrangements (3 dummy variables)

Patient history
Employment: 9 dummy variables for each of the years *t-1*, *t-2*, and *t-3*, (27 dummy variables in total). For each year, the variables represents combination of an income-based employment status measure and information about regular work-hours.

Earnings: Income measured in NOK (inflated to 2005-value) for each of the years *t-1*, *t-2*, and *t-3*

Absenteeism: One dummy for each of six possible sick leave histories last three years (6 dummy variables)

Neighborhood
One dummy variable for each neighborhood in Norway (12921 dummy variables)

Calendar time
One dummy variable for each starting month occurring in our observation window (56 dummy variables)

Physician characteristics (see description below)
Age (5 dummy variables based on 10-year grouping)
Specialist education (2 dummy variables)
Gender (2 dummy variables)
Sharing office with other physicians (2 dummy variables)
Fixed or variable salary (2 dummy variables)
Taking part in emergency service (2 dummy variables)
Number of patients on capitation list
Desired number of patients relative to actual number
Two scalar indicators for physician leniency (see below)
One scalar indicator for treatment quality (see below)

Derivation of the physicians' practice style indicators

The physicians' practice style indicators have been computed by means of auxiliary regressions. Leniency is measured by the physicians' propensity to issue absence certificates to the employees on their capitation list. To compute physician leniency, we exploit

data on all Norwegian employees (not only those with long-term absence), and set up auxiliary linear probability models in exactly the same fashion as we did to compute grading propensity, with a dummy variable for each physician along with a large number of individual controls. We compute two leniency-indicators, one based on the annual incidence of certified sick leave (denoted overall leniency) and one based on the probability of having a long-term sick leave (denoted long-term leniency). The dataset includes more than 7 million worker-year observations. The estimated coefficients attached to the physician dummies are taken to represent physician leniency, and subsequently used as additional control variables in Equations (3) and (4) along with other observed physician characteristics. Again we have estimated separate models for men and women. We have then used the leniency indicators estimated on female data as additional controls in the male instrumental variables model and vice versa.

Physicians' treatment quality is clearly a latent variable, and since the health status of patients is also unobserved, it is not obvious how we should control for this variation. We may hypothesize that the quality of a physician's medical advice is correlated to the patients' survival, conditioned on all observed patient characteristics, in which case we may use data on mortality to compute a proxy for physician influences beyond their absence certification practices. We realize, of course, that physicians have limited influence on their elderly patients' survival, and, hence, that a practice style measure based on observed mortality is likely to be very noisy. To compute our indicator, we have exploited an additional cut of our data consisting of elderly retirees (around 4 million person-year observations). Again, we have used an auxiliary linear probability model, this time with survival as the outcome and again with physician dummies and patient charac-

teristics as explanatory variables. Since the risk set of elderly retirees do not contain any of the employees in our long-term absence dataset, we have estimated a common indicator for men and women this time, and used it along with the other practice style indicators in the instrumental variables models.

References

Abbring, J. H. and Van den Berg, G. J. (2003) The Non-Parametric Identification of Treatment Effects in Duration Models. *Econometrica*, Vol. 71, No. 5 (September), 1491-1517.

Andrén, D. and Andrén, T. (2008) Part-Time Sick Leave as a Treatment Method? HEDG Working Paper 08/011. University of York.

Andrén, D. and Andrén, T. (2009) How to Evaluate the Impact of Part-Time Sick Leave on the Probability of Recovering. Working Paper 13/2009. Örebro University.

Bratsberg, B., Fevang, E., and Røed, K. (2010) Disability in the Welfare State: An Unemployment Problem in Disguise? IZA Discussion Paper No. 4897.

Franche, R-L, Cullen, K., Clarke, J, Irvin, E., Sinclair, S, Frank, J and The Institute for Work & Health Workplace-Based RTW Intervention Literature Review Research Team (2005) Workplace-Based Return-to-Work Interventions: A Systematic Review of the Quantitative Literature. *Journal of Occupational Rehabilitation*, Vol. 15, No. 4, 607-631.

Grytten, J. and Sørensen, R. (2003) Practice Variation and Physician-Specific Effects.

Journal of Health Economics, Vol. 22, 403-418.

Hagen, E. M., Grasdal, A., Eriksen, H. R. (2003). Does Early Intervention with a Light Mobilization Program Reduce Long-term Sick Leave for Low Back Pain: A 3-year Follow-up Study. *Spine*, Vol. 28, 2309-2316.

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

Høgelund, J., Holm, A., McIntosh, J. (2010) Does Graded Return-to-Work Improve Sick-Listed Workers' Chance of Returning to Regular Working Hours? *Journal of Health Economics*, Vol. 29, 158-169.

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

Kausto, J., Miranda, H., Martimo, K-P, Viikari-Juntura, E. (2008) Partial Sick Leave – Review of its Use, Effects and Feasibility in the Nordic Countries. *Scand J Work Environ Health*, Vol. 34, No. 4, 239-249.

Malmivaara, A., Häkkinen, U., Aro, T., Heinrichs, M-L, Kosenniemi, L., Kuosma, E., Lappi, S., Paloheimo, R., Servo, C., Vaaranen, V., Hernberg, S. (1995) The Treatment of Acute Low Back Pain –Bed Rest, Exercises, or Ordinary Activity? *The New England Journal of Medicine*, Vol 332, No. 6, 351-355.

Manski, C. F. (1993) Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, Vol. 60, No. 3, 531-542

Markussen, S. (2009a) The Effects of Sick-Leaves on Earnings. Memorandum No.

20/2009, Department of Economics, University of Oslo.

Markussen, S. (2009b) Closing the Gates? Evidence from a Natural Experiment on Physicians' Sickness Certification. Memorandum No. 19/2009, Department of Economics, University of Oslo.

Mykletun, A., Eriksen, H. R., Røed, K., Schmidt, G., Fosse, A., Damberg, G., Christiansen, E. C., Guldvog, B. (2010) *Tiltak for reduksjon i sykefravær: Aktiviserings - og nærværsreform*. Ekspertgrupperapport til Arbeidsdepartementet. English Summary available at: http://www.regjeringen.no/en/dep/aid/topics/welfare-policy/inclusive-working-life/summary-of-.html?id=592754

OECD (2009) Sickness, Disability and Work - Addressing Policy Challenges in OECD Countries. High-Level Forum Stockholm, 14-15 May 2009. OECD, Paris.

Sieurin, L., Josephson, M., and Vingård, E. (2009) Positive and negative consequences of sick leave for the individual, with special focus on part-time sick leave. *Scandinavian Journal of Public Health*, Vol. 37, 50-56.

Waddel, G. (2004) *The Back Pain Revolution* (Second Edition). Elsevier Health Sciences.

Waddell, G. and Burton A. K. (2006). *Is Work Good for Your Health and Well-being?* Report for Department for Work and Pensions. TSO (The Stationery Office), London.