CAN COMPULSORY DIALOGUES NUDGE SICK-LISTED WORKERS BACK TO WORK?*

Simen Markussen, Knut Røed and Ragnhild C. Schreiner

We evaluate the impacts of a compulsory dialogue meeting for long-term sick-listed workers in Norway. The meeting is organised by the local social security administration after around six months of absence; its purpose is to bring together the absentee, the employer and the family physician to discuss whether arrangements can be made to facilitate partial or full work resumption. Our causal analysis is based on random-assignment-like geographical variation in the meeting propensity. We find that the meetings reduce absence duration considerably, both through a notification and an attendance effect. Moreover, estimated benefits by far exceed estimated costs.

Programmes insuring workers against income losses during sickness absence are of major importance in many welfare states. These programmes are typically highly valued by workers as well as voters. Yet, researchers have found evidence of serious moral hazard problems: the higher and more long-lasting is the sick pay, the higher is the rate of sickness absence (Johansson and Palme, 1996; Henrekson and Persson, 2004). Higher absenteeism may in turn imply a greater risk that some absence spells develop into more serious long-term disabilities with low likelihood of successful return to regular employment. Policy makers therefore face a difficult trade-off and are (or should be) interested in measures that can reduce the moral hazard problems associated with sick-leave insurance, such that the desired level of insurance can be provided at the lowest possible cost. This article examines the effects of one such policy, namely a compulsory dialogue meeting (DM) for long-term sick-listed workers in Norway.

Like in most other industrialised countries, Norway's sick-leave insurance system entails moral hazard problems with respect to the behaviours of both employers and employees. The employee receives a 100% wage replacement (up to a relatively high ceiling; see details in next Section) from the first day of absence and for up to one year. The cost the first 16 days of each absence spell is fully covered by the employer, after which the public purse takes over the bill. There is no experience rating. This insurance design (which with some variations is typical for European OECD countries (OECD, 2010) may imply that employees exert less effort to avoid and escape from absenteeism than what is optimal from a social planner's point of view, whereas the employers exert too little effort to facilitate work resumption and prevent the employee from losing the foothold in the labour market. One way to reduce the moral hazard problems is to monitor the behaviour of employers and employees. The main device used for this purpose is a requirement that absence spells exceeding three days

* Corresponding author: Ragnhild C. Schreiner, Ragnar Frisch Centre for Economic Research, Gaustadalléen 21, 0349 Oslo, Norway. Email: r.c.schreiner@frisch.uio.no.

[1]

(eight days in some firms) are certified by an authorised physician.[1] However, it is well known that physicians may be rather poor gatekeepers, not only because their health assessments largely are based on what the patients choose to tell them but also because they have financial incentives to ensure that their patients are sufficiently satisfied to not find a new physician (Markussen *et al.*, 2013; Markussen and Røed, 2016). Thus, additional monitoring mechanisms may be required. Given that the financial cost of continuing a sick-pay spell a few extra days may be negligible for both the absentee and the employer, it is possible that even very small hurdles, for example, the need for an unpleasant conversation between the two of them, are sufficient to postpone work resumption repeatedly. Procrastination in intertemporal effort choices is widespread and a growing empirical literature indicates that many individuals discount the future in a hyperbolic fashion; that is, with a bias towards the present (DellaVigna and Paserman, 2005; Paserman, 2008; Cockx *et al.*, 2014). This implies that activities for which future benefits must be weighed against immediate costs tend to be postponed repeatedly, even when the activities are optimal from a long-term perspective. Hyperbolic employers and employees will always be tempted to delay unpleasant work resumption efforts yet another few days. In such cases a small 'nudge' may be all that is needed to speed things up. There is a parallel here to the literature on policy interventions for unemployed workers, where positive effects have been reported in relation to relatively small interventions, such as summoning the job seekers to meetings; see, for example, Dolton and O'Neill (1996), Gorter and Kalb (1996) and Hägglund (2011).

The meeting evaluated in this article may be viewed as a combined nudging and monitoring device. It is organised by the local social security administration (SSA) around six months into the sick-leave spell and its main purpose is to bring together the absentee, the employer and the certifier of the sick leave to discuss whether arrangements can be made at the workplace that make full or partial work resumption possible. Examples of such measures can be to alter the number or nature of tasks at work, to implement a home office, or to adjust work hours. An important feature of the Norwegian sick-pay insurance system is that it allows partial (graded) absence, implying, for example, that a sick-listed employee works at a 50% capacity (and receives the normal wage from the employer for this part), while collecting 50% sick-pay benefits. Previous evidence has indicated that promoting partial rather than full absence is a fruitful strategy towards reducing overall absenteeism (Markussen *et al.*, 2012) and the sick-pay legislation actually requires physicians to issue partial absence certificates for all absences exceeding eight weeks, unless there are strong medical grounds for maintaining 100% absence. Yet, the majority of absence spells are not graded, and the DM thus provides an arena where the social security administration can push for partial as well as full work resumption.

The questions we seek to answer in this article are whether or not the DM has achieved its aim of speeding up the process of partial or full work resumption, and – if so – whether it has done so in a cost-efficient way. In addition, we are interested in identifying the mechanisms behind any DM effects. More concretely, we distinguish

---

[1] If the employer is part of the 'Inclusive Workplace Agreement' (IA), workers can self-certify absence lasting up to eight days. This is the case for approximately 50% of Norwegian employees.

between 'notification effects' and 'attendance effects'. The DM is typically convened three weeks before it is supposed to take place through letters from the social security administration to the sick-listed employee, the employer and (if deemed appropriate by the caseworker) the family doctor (or the sick-leave certifying physician if this is different from the family doctor). If the sick leave ends after the letters have been sent, but before the scheduled time of the meeting, the meeting will normally be cancelled. The notification effect encompasses the responses triggered by the summons letters before any meeting actually takes place. It bears some resemblance to the notorious 'threat effect' frequently encountered in the unemployment insurance literature with reference to the effect of being summoned to an activation programme (Black *et al.*, 2003; Geerdsen, 2006; Rosholm and Svarer, 2008; Graversen and Larsen, 2013). These effects may be present even when the programme itself is modest and thus has a nudging character. Hägglund (2011), for example, finds positive pre-programme effects related to a series of meetings intended to support and monitor job-search activities. In our case, however, the notification effect not only refers to the behaviour of the insured worker but also to the behaviours of the employer and/or the physician. The attendance effect encompasses the impacts following from the meeting itself, for example, in terms of an agreed strategy for work resumption involving modified tasks and/or work hours.

The DM was introduced in Norway in 2007 but there are no data on meetings actually held until 2009. Our analysis therefore includes spells starting between January 2009 and December 2010; it is based on complete administrative registers for all physician-certified absence spells in this period. To identify causal impacts of the DM, we exploit the fact that even though the meeting is in principal compulsory, there has been ample scope for local social security administrations to make exemptions. This has resulted in a considerable geographical variation in both the overall use of DMs and in their precise timings within absence spells. We argue that from the workers' and firms' points of view, this gives rise to a random-assignment-like variation in the duration-specific probabilities of being called to a DM, which makes it possible to identify their causal effects. While our data contain precise information about the timing of all realised meetings, we have no direct information about the calls. However, as we explain in more detail below we can use information about realised meetings, combined with prior information about the typical duration from the call to DM realisation, to distinguish notification from attendance effects.

We examine the effects of the DM on sick-leave duration within a mixed proportional hazard rate framework where we aim at distinguishing notification from attendance effects by including in the return-to-work hazard one variable representing the probability of being called to a meeting and another representing the probability of having already (and recently) participated. We estimate two different models, one where we treat any form of work resumption (partial or full) as the outcome of interest, and another where we treat full work resumption as the outcome of interest while disregarding partial work resumption. Furthermore, we use data from a period before the DMs were implemented in a placebo regression.

Our study relates to Johansson and Lindahl (2013), who examine the effects of an information meeting (IM) on the duration of absence spells of (largely) non-employed sick-pay claimants; that is, persons who have already lost the job that originally made

them eligible for sickness benefits. The IMs are organised by Swedish local SSA offices and the purpose of the meeting is to inform absentees about the criteria for continued sickness benefits. The study is based on data where the timing of the call to the IM is randomised. The results from the evaluation of the IMs suggest a significant positive effect on the exit hazard from sick leave. However, for these persons, the alternative to sick-pay will often be unemployment insurance, and the study indeed shows that the increase in outflow from sick leave is partly met by an increase in the inflow to registered unemployment. By contrast, our own study focuses on employed sick-listed workers where the alternative to continued absence is partial or full work resumption. The main contribution of the article is twofold. First, to our knowledge, it is the first article to evaluate this type of nudging-like policy to induce sick-listed workers to return to work. Second, we propose a novel framework for identifying the effects of being notified of a treatment in situations where only actual participation in treatment (and not the notification) is observed and there is a self-selection into treatment caused by the notification.

Our main finding is that the DM has a positive and substantial effect on the hazard to partial as well as full work resumption. Both notification and attendance effects contribute to earlier work resumption and, based on a simulation exercise, we find that the two effects are of roughly the same quantitative importance. Together, the notification and attendance effects estimated in this article imply that for each realised DM, the duration until partial or full work resumption is reduced by approximately 20 calendar days (including weekends). Given the low cost of arranging the meetings, we also find that the meetings are highly cost effective, in the sense that the value of the additional work they generate – as measured by earnings (which are again equal to the saved insurance payouts) is much larger than the meeting costs.

1. Data and Institutional Setting

Conditional on receiving a certificate from a physician, all absentees are entitled to a 100% wage replacement ratio for up to one year. The replacement ratio is 100% up to a ceiling of six times the base amount in the Norwegian pension system. The base amount is adjusted every year and was equal to 85,245 NOK in 2013. Using the average exchange rate for 2013, the base amount corresponds to approximately US$14,500.[2] A major challenge with the Norwegian system is the lack of economic incentives for absentees to return to work during the first year on sick-leave benefits, as well as the weak incentives for the employer to promote work resumption. The employer pays for the first 16 days of the absence spell whereas the social security administration (SSA) covers the benefit payments thereafter. This means that the social security system undermines the employer's economic incentives to exert effort in helping long-term absentees back to work. Indeed, there is empirical evidence that employers sometimes discourage long-term absentees from returning to work due to the risk of the absentee starting a new spell shortly thereafter (Fevang *et al.*, 2014). Moreover, if the absence

spell outlast the one-year sick-pay period, a separate employment protection for absent workers no longer applies, implying that the employer can legally fire the worker without having any responsibility for subsequent social insurance payments.

Several measures, the DM included, have been implemented to counteract these incentives. Their purpose is both to facilitate the return to work and to monitor that the absentees reveal their actual need for absence. For employed workers, the employer is obliged to arrange a meeting with the employee during the first seven weeks of the spell to agree on a plan for work resumption, if necessary involving changes in tasks or work hours. No later than 26 weeks into the spell, the SSA is supposed to organise a DM. The meeting was introduced in 2007 to induce long-term absentees to return to work fully or partly, as opposed to continuing on a path to disability insurance dependency. The DM is intended to provide a setting where the employer and the absentee can discuss possible measures to be made at the workplace to make work resumption possible. The certifier of the sick leave should attend the meeting if the caseworker at the SSA office considers it to be appropriate. Even though the local SSA offices are required by the law to arrange a DM within 26 weeks of the absence spell, there is significant variation between counties in the frequency and timing of the DM. There are 19 counties in Norway with an average of 24 SSA offices within each county. The county administrations face binding budgets and this can result in considerable differences in the extent of use of different labour market programmes, including the DM. Failing to organise a DM can be justified as an exemption, something the law actually allows room for. The law states that an exemption is allowed if 'such a meeting is assumed to be clearly unnecessary' (Folketrygdloven § 8-7,8 a).[3] As it turns out, this 'exemption option' has been interpreted and exploited diversely in different parts of the country, partly due to differences in local administrators' views about the usefulness of the meetings, and partly due to differences in the pressure on caseworkers caused by other tasks, for example, related to social assistance, unemployment, or disability insurance claims. At a typical SSA office, only 4% of the caseworker resources are spent on tasks directly related to sick leave (Väageng *et al.*, 2015); hence even small differences in, say, the workloads related to unemployment or disability insurance claims, may have large consequences for the availability of resources required to follow up sick-listed workers. From the sick-listed workers' point of view, this potentially constitutes a variation in the likelihood of being exposed to – and/or in the precise timing of being called to – a DM, which is as good as randomly assigned.

Figure 1 illustrates the variation across counties in DM exemption practices, as reflected in the share of all long-term absence spells (longer than five months) started between January 2009 and December 2010 for which a realised DM is recorded. The share varies from around 13–18% in Østfold and Oslo (the capital) to more than 30% in Vestfold and Telemark.

We have data on the exact timing on all realised DM meetings, however we do not observe the exact timing of the calls, and whether an absentee is called to a meeting and then returns to work before the meeting is realised. Further, we observe the duration of

---

[3] The most common reasons for making exemptions are if the absentee is severely ill and/or admitted to a health institution, if the absentee is expected to return to work shortly after the time of the DM, or is part of an active labour market programme that is likely to lead to an upcoming return to work.
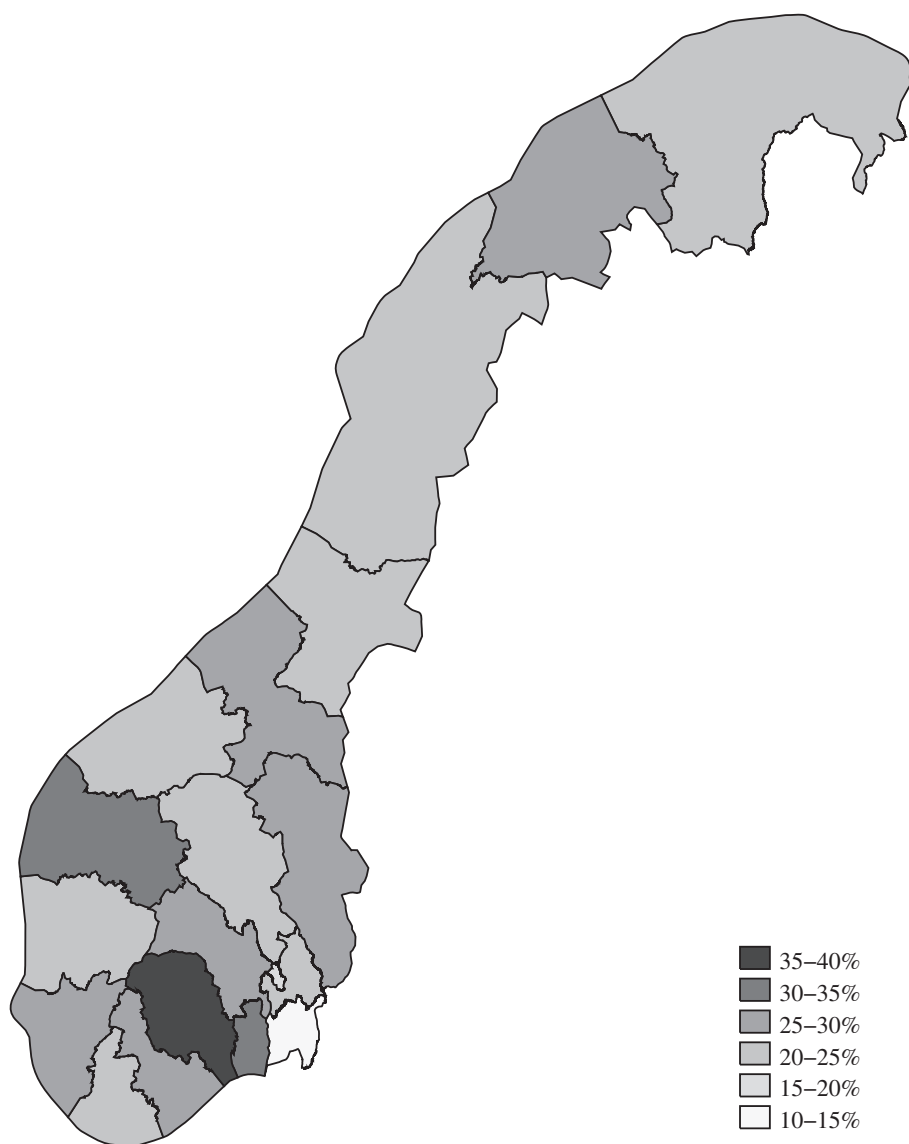
■	35–40%
▨	30–35%
▨	25–30%
▨	20–25%
▨	15–20%
□	10–15%

Fig. 1. *Geographical Variation in the Use of DMs*

Notes. Map of Norway, where the 19 counties are divided into six groups by the share of long-term spells (longer than five months) started between January 2009 and December 2010 for which a realised DM is recorded. A darker shade illustrates a more intensive use of DMs.

sick-leave spells and transitions to and from full or partial work. The main data used in the empirical analysis consists of all physician-certified sick-leave spells in Norway, unconditional on spell duration, that were started between January 2009 and December 2010. As pointed out above, The DM was implemented already in 2007; however, we do not have data on the use of the meeting before 2009. This also implies that in order to find a time period without DMs to use as a suitable control period for a set of placebo estimations, we

must go back to spells that started before the summer of 2006 and thus were unlikely to be affected by the introduction of DMs in 2007. Moreover, to avoid interference from another reform in July 2004 (which changed the regulations regarding sick-leave certification), we use spells that started between July 2004 and June 2006 to establish a control group of pre-DM spells. Inclusion in the data is further conditional on full sick-pay eligibility, meaning that the absentee cannot have received any sickness benefits over the previous six months. Finally, we require that the absentee is between the age of 18 and 66, is registered in the Employee Register (*Arbeidstakerregisteret*), and have an annual income exceeding an amount corresponding to approximately US$14,500. All absentees are followed from the beginning of the spell until a transition to resume partial or full work. A spell is right censored if no transition takes place within the twelve-month period of sickness benefits.[4]

Table 1 shows descriptive statistics on the data used in the empirical analysis. We estimate two alternative hazard rate models with different definitions of the work

Table 1

Descriptive Statistics

	Full-time sick-leave spells		All sick-leave spells	
	DM period (1)	Pre-DM period (2)	DM period (3)	Pre-DM period (4)
Fraction of females (%)	57.3	56.3	58.1	57.0
Mean age at entry to sick leave	40.0	39.9	40.1	40.0
Level of achieved education				
High school (%)	44.8	48.1	44.4	47.6
University/college or higher (%)	29.1	27.0	30.1	27.8
Immigrant background (%)	20.8	15.7	20.5	15.5
Mean income in base amounts	5.0	5.0	5.0	5.0
(US$14,500)	(2.2)	(2.1)	(2.2)	(2.1)
Mean duration until partial or full	29.5	28.2	N/A	N/A
work resumption (days)	(62.0)	(59.1)		
Mean duration until full work	N/A	N/A	39.6	37.4
resumption (days)			(78.0)	(75.0)
Fraction with a DM (%)	1.3	N/A	2.4	N/A
Mean time before DM (in days)	182	N/A	184	N/A
Fraction returning to work (%)	98.5	98.5	97.9	97.9
Number of spells	2,014,576	1,880,895	2,115,424	1,966,940
Number of individuals	1,060,256	1,009,139	1,098,709	1,042,308
Long-term spells > 26 weeks				
Fraction with a DM (%)	33.6	N/A	36.5	N/A
Fraction returning to work (%)	65.6	64.4	70.5	68.0
Number of spells	74,245	61,732	132,975	112,025

Notes. Standard deviations are given in parentheses. Column (1) shows all sick-leave spells started between January 2009 and December 2010. Column (2) shows all sick-leave spells started between July 2004 and June 2006. Columns (3) and (4) are a sub-sample of columns (1) and (2) and shows all full-time (100% sick leave from work) spells started within this period.

[4] Disability benefits are not relevant before employed workers have exhausted their rights to sickness benefits (the latter benefit is more generous). Since the absentees in our data have full sick-pay eligibility, disability benefits play no role in the analysis on short term effects of DMs.

resumption event and corresponding analysis populations. Since both models are estimated on data for both the pre-DM and DM periods, we have in total four data sets used for estimation. Columns (1) and (2) show descriptive statistics for all sick-leave spells that were classified as full-time (100%) absence from the beginning, for the periods with and without DMs respectively. These spells are used to examine DM effects on any degree of work resumption – partial or full. Most of the absence spells in Norway (95%) are indeed full-time at the time of entry and, as explained above, an important aim of the DMs is to encourage (at least) a gradual return to work for those who have little or no contact with their employer during the spell. Columns (3) and (4) show the corresponding descriptive statistics for all spells – including those that start out with a graded absence certificate and, hence, where the claimants already partially work from the start of the absence spells. These spells will be used to examine DM effects on the transition to full work resumption (and thus exit from sick leave) only. In each of the four data sets, there are around two million absence spells – experienced by about one million persons. For the spells starting with full-time absence, the average duration until partial or full work resumption takes place is 28–29 days and around 3–4% of the claimants have not resumed any kind of work after six months (the typical timing of a DM). For all spells (including those that start out with partial work), the average duration until full work resumption is close to 40 days, and around 6% of the claimants are still on sick-pay after six months.

2. Empirical Analysis

In this Section, we set up and estimate statistical models designed to identify the causal effects of DMs on work resumption. The basic idea of our identification strategy is to exploit the random-assignment-like variation in absentees' exposure to DMs generated by differences in DM intensities across counties. Hence, the first step of our empirical approach is to compute county-specific DM intensities for each duration-week (subsection 2.1). We then exploit the duration-specific spatial variation in these intensities to identify and estimate the causal effects of being called to and of participating in a DM (subsection 2.2). Finally, we discuss potential challenges towards our identification strategy and assess the trustworthiness of the alleged causal interpretation of our estimated parameters (subsection 2.3). Since we exploit the county variation in the use of DMs to identify causality, our estimates are based on absentees who potentially are treated differently in different counties. To the extent that the true causal effects are heterogeneous, this implies that our estimates will be representative for this 'marginal' population. This is also the population that primarily will be affected by attempts to scale up or down the use of DMs. Estimates based on observed county variation is thus arguably policy relevant. The main aim of our empirical analysis is to examine the extent to which DMs influence the propensity to partially or fully resume work and, further, to decompose any estimated impacts into notification and attendance effects. Intuitively, it appears possible to separate notification from attendance effects on the basis of their timing within spells. Effects occurring prior to the typical timing of the meeting reflect notification, while effects occurring afterwards reflect attendance. However, if there really is a notification effect, this may itself imply that the population actually attending the meeting is selected. This

particular selection problem remains even if the calls to the meetings are completely randomly assigned. Hence, to distinguish the two effects correctly, we need to model explicitly how any shifts in hazard rates following from notification changes the population at risk of being exposed to an attendance effect also. We do that by means of a hazard rate model, allowing for observed as well as unobserved heterogeneity.

2.1. *Measuring County-specific Meeting Intensity Profiles*

To compute county-and-duration-specific DM intensities relevant to each absence spell *s*, we use all absence spells exceeding 16 week duration and set up a transition rate model with participation at a DM as the endogenous event (shorter spells are dropped for the reason that a DM almost never occurs during the first 16 weeks). The model consists of one observation per absentee per week of the absence spell until the absentee is no longer under risk of attending a meeting. This means that an additional observation is included for each week the absentee has not attended a meeting or terminated the spell. The spell is right censored when the absentee returns to work, or after 12 months.

This computation strategy is similar to that of Markussen and Røed (2014), who study the effects of a variety of vocational rehabilitation programmes by exploiting local variation in the use of the different programmes as a source of randomness in the probability of being assigned to the different programmes. When constructing the measures of DM intensities, the contribution of a person's own spell must be removed from the indicator to avoid a mechanical correlation between the indicator and anything unobserved about the spell. To simplify these calculations, we follow Markussen and Røed (2014) and use a linear transition rate model to construct the treatment intensities. Let DM_{sjd} be the event of attending a meeting for an absentee with spell *s* who has been under risk of meeting attendance for $d > 16$ weeks in county *j*:

$$DM_{sjd} = \boldsymbol{\delta}_d + \boldsymbol{x}'_{st}\boldsymbol{\theta} + u_{sjd}, \tag{1}$$

where $\boldsymbol{\delta}_d$ is a vector of weekly duration dummy parameters and $\boldsymbol{x}_{st}$ is a vector of individual observable control variables including age, sex, medical diagnosis, earnings level, industry of employment, education (level and type), contracted work-hours, local labour market tightness, and calendar time. To avoid invalid functional form restrictions, all the controls are included in the form of relatively large numbers of indicator variables; see Appendix B for details. The measure of treatment intensity faced by the individual with spell *s*, still at meeting risk at duration *d* in county *j* is constructed by calculating the mean residual within county for each duration, while removing the contribution of spell *s*:

$$\phi_{jd,-s} = 1/(N_{jd} - 1)\left(\sum_{k \in N_{jd}} \hat{u}_{kjd} - \hat{u}_{sjd}\right), \tag{2}$$

where N_{jd} is the number of spells still at meeting risk at duration *d* in county *j*, and $\hat{u}_{sjd}$ is the residual from regression (1). As we are aiming at disentangling notification and attendance effects of the DMs, we are interested in both the probability of attending a

DM in each given duration week and in the probability of already having attended a meeting. Both variables are expressed in terms of the county and duration-specific treatment intensities, which are scaled in order to make the attendance probabilities vary in the range of zero-one. For spells, we hence express the probability of being called to a DM in week d conditional on not having been called before (the weakly hazard) as follows:

$$h_{sjd} = \begin{cases} 0 & \text{for } d = 0, 1, \ldots, 16 \\ \hat{\phi}_{jd,-s} + \hat{\delta}_d + c & \text{otherwise,} \end{cases} \tag{3}$$

where $\hat{\delta}_d$ is the estimated coefficient on element d of the vector of duration parameters from (1), and c is a constant such that $c + \hat{\delta}_{d=26} = D\bar{M}_{d=26}$, where $D\bar{M}_{d=26}$ is the observed mean DM transition rate for all spells in Norway at duration week 26.

From this expression, we can derive the duration-specific probability of already having attended a meeting, S_{sjd} as:

$$S_{sjd} = S_{sj,d-1} + (1 - S_{sj,d-1})h_{sjd}, \tag{4}$$

with $S_{sj,d=0} = 0$. The second part of this expression is the unconditional probability of attending a DM for each duration week which we denote g_{sjd}:

$$g_{sjd} = (1 - S_{sj,d-1})h_{sjd}. \tag{5}$$

Figure 2 illustrates the spatial variation in the two predicted attendance probabilities (attending this week (g_{sjd}) and already having attended (S_{sjd})) over spell duration, by comparing two counties in each end of the meeting intensity distribution with the national average. The filled circles in panels (a) and (b) are
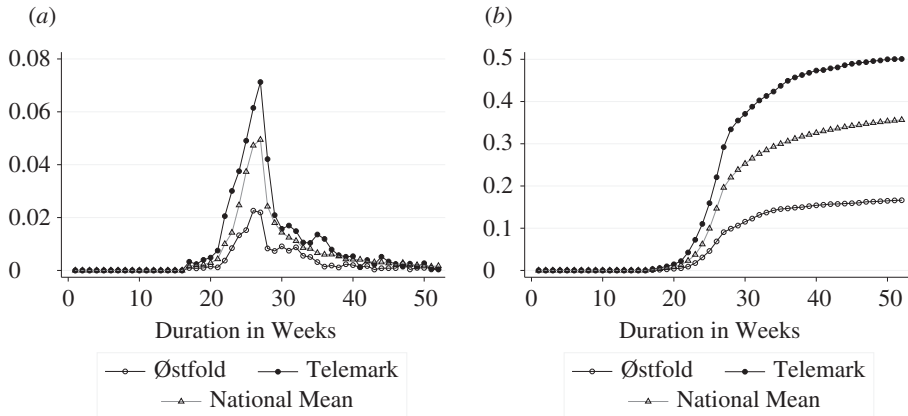

Fig. 2. *DM Attendance Probabilities. (a) Probability of Attendance This Week. (b) Probability of Already Having Attended*
Notes. Panel (a) illustrates the variation in the predicted probability of attending a DM over spell durations for the county in Norway that tends to use the DM the least (Østfold; illustrated by hollow circles), the county that tends to use the DM the most (Telemark; illustrated by filled circles) and the country as a whole (illustrated by triangles). Panel (b) illustrates the variation in the predicted probability of already having attended a DM, also for Østfold, Telemark and the whole country.

the duration-specific means of the attendance probabilities in Telemark, the county that tends to use DMs the most, while the hollow circles are the duration-specific means in Østfold, the county that tends to use the DMs the least. Finally, the triangles represent the duration-specific national means. As can be seen from the Figures, the attendance probabilities vary considerably both over spell duration and over counties. This allows us to condition on county and spell duration and identify the effects of notification and attendance from variation in meeting intensities over spell duration within each county. In line with national regulations, there is a clear attendance spike around week 26, yet a number of meetings are held sometime before or after week 26 also. Conditional on sick-pay exhaustion (52 week duration), the probability of having attended a DM at some stage during the spell varies from around 18–50% over counties, with a national mean of approximately 35%.

2.2. *Effects on the Work Resumption Propensity*

We model the probability of resuming work – partially or fully – by means of a mixed proportional continuous time hazard rate model (MPH) with piecewise constant duration effects estimated separately for each week. Given the suspected selectivity of DM participation – both due to selectivity in the calls to the meetings and due to the selectivity resulting from the possible effects of the calls – we do not exploit data on actual DM participation at all in this subsection. Instead, our key explanatory variables are going to be proxy-variables representing the probability of being notified about a forthcoming DM and the probability of already (and recently) having participated. Although it is clear from the previous discussion that we intend to derive these probabilities on the basis of the county-by-duration-specific DM intensities, it is far from obvious exactly how this should be done. Since DM notifications are unobserved, we need to make assumptions regarding their timing relative to the actually held meetings, and also regarding the duration of any notification effects. In addition, we need to make assumptions about the duration profile of any attendance effects. Since none of these assumptions can be completely based on prior knowledge, a central element of our modelling strategy is to test out a number of (reasonable) alternative specifications and then to choose the best one based on an information criterion. For the moment, we take an agnostic view on this and include in the hazard rate the unspecified variables N_{sjd} and A_{sjd} as representing notification and attendance, respectively, together with controls for individual (and spell) heterogeneity, calendar time, spell duration and county. Let x'_{st} be a vector of observable characteristics, γ_j be a vector of county fixed effects, λ_t denote calendar month fixed effects (monthly dummy parameters), ρ_d denote weekly spell duration parameters and v_i be a measure of individual-specific unobserved heterogeneity. The work resumption hazard rate can then be expressed as:

$$\theta_{sijtd} = \exp(x'_{st}\alpha + \lambda_t + \rho_d + \gamma_j + \beta^1 A_{sjd} + \beta^2 N_{sjd} + v_i), \qquad (6)$$

where β^1 and β^2 are parameters capturing the attendance and notification effects, respectively. The vector of observed characteristics x_{st} is exactly the same as the one we

used to compute the county-specific DM indices in (1); see Appendix B for details.[5] It may be noted that while observed individual explanatory variables are spell-specific and time-varying, the individual unobserved covariate, v_i, is person specific (hence the i subscript) and time invariant. This implies that we exploit the existence of repeat spells to disentangle the impacts of duration dependence and unobserved heterogeneity; see, for example, Van den Berg (2001). Person-specific unobserved heterogeneity can in our case be justified by the fact that we have controlled for the spell-specific medical diagnoses; hence the role of the unobserved covariate is primarily to capture more deep-rooted individual characteristics, such as motivation, work ethics and general health status.

An important point to note from this specification is that since we have included county and duration fixed effects in the hazard rate, systematic differences in work resumption rates between counties or over spell duration will not contribute to identification of the causal parameters of interest. The two DM effects are instead identified by shifts in the hazard rate caused by the county-specific meeting intensities at durations corresponding to the typical timing of notification and recent participation, respectively, namely, by the variation in duration profiles across counties and their correlations with the DM intensity variables. The model is estimated with non-parametric maximum likelihood and the number of support points in the (discrete) heterogeneity distribution is chosen on the basis of the Akaike information criterion (AIC); see Appendix A for details.

In order to carry out this estimation strategy, we first need to specify the two key variables N_{sjd} and A_{sjd}. The notification variable N_{sjd} should ideally represent the county-by-duration-specific probability of receiving a call to a DM. While we have a good indicator of the probability of weekly participation, we have no data on meeting calls. We have been informed by the Norwegian social security administration that a commonly used practice is to convene the DM three weeks prior to its planned date. A natural assumption to make is thus that the county-by-duration-specific probability of receiving a meeting call is approximately proportional to the corresponding probability of participation three weeks later.[6] The

---

[5] In the model presented in this Section, we have controlled for medical diagnosis by means of 31 dummy variables based on average spell duration (from older data) measured in five-days intervals up to a duration of 150 days and then one category for diagnoses with an average spell durations exceeding 150 days. We have also estimated the model with two alternative specifications of diagnosis controls, one where the set of dummy variables is extended to include separate categories for each 5-day interval of average diagnosis duration exceeding 150 days and another version where we included a separate dummy for each of the 200 most common diagnoses. None of these alternative specifications changed the results presented below to any noticeable extent (the first model implied, however, that we had to drop a number of observations due to multicollinearity problems, and the latter specification implied dropping spells with less common diagnoses).

[6] Note that the relationship will not generally be proportional, as the ratio of the call-propensity to the realised meeting propensity depends on the level of the involved baseline hazard rates, on the county-specific probability of participation, as well as on the size of the notification effect. For example, in the simplified case where the meeting is held one week after notification, it can be shown that the functional form relationship between the participation and notification probabilities can be expressed as

$$N_{sjd} = [\exp(-h)]/\{\exp[-[\exp(\beta^2)h] + g_{sj(d+1)}(\exp(-h) - \exp\{-[\exp(\beta^2)h]\})]\}g_{sj(d+1)},$$

where h is the hazard rate in the week of question in the absence of notification. Based on the results presented later in this article, we infer from this formula that the call propensity is approximately 7–8% higher than the realised meeting propensity in the model with full work resumption as the outcome of interest and 13–14% higher in the model with both partial and full work resumption. The largest deviations occur in the counties with the lowest call propensity. However, attempts to adjust for these differences indicate that although this measurement error causes $\exp(\beta^2)$ to be overestimated by 7–8% and 13–14% in the two models respectively, it has negligible effects on our simulation-based interpretation of the results.

timing of notification is likely to vary somewhat across spells, however, and in order to assess this further, we specify a more general measure by leading the attendance probabilities derived in Equation (5) by z_N weeks:

$$N_{sjd} = g_{sj(d+z_N)}. \tag{7}$$

This specification of the notification variable implies that the work resumption hazard during a particular week is modelled as a function of the county-specific participation rate applying z_N weeks later. The specification further involves an assumption of the notification effect lasting for exactly one week. This assumption is made since the data do not allow us to distinguish between a one-week notification effect and a more dispersed effect pattern over the weeks close in time to the actual week of notification. Hence, the coefficient attached to N_{sjd} needs to be interpreted with some care. It is also of some interest to examine specifications of N_{sjd}, based on alternative choices of z_N. For example, specifying the lead as two instead of three weeks may be preferable if either the call tends to be submitted somewhat closer to the actual meeting, or if the main effect comes during the week after the notification was received. Similar arguments can be made regarding the specification of the attendance variable. We will define the attendance variable A_{sjd} such that it measures the county-by-duration-specific probability of recently having participated at a DM. In this case, we need to operationalise the concept of 'recently'; that is, the period of time following a DM for which we allow for a causal effect. Assuming that the attendance effect lasts for z_A weeks, we specify the attendance variable as the probability of having participated at a DM sometime during the last z_A weeks, which based on the notation in (4) can be expressed as:

$$A_{sjd} = S_{sjd} - S_{sj(d-z_A)}. \tag{8}$$

The specification of a constant effect lasting for a given number of weeks after meeting attendance is of course somewhat arbitrary and it should be interpreted as a simplification rather than as a substantive assumption. In practice, we expect there to be a dynamic effect pattern, possibly with effects that taper off with time. The probability of having participated during the last z_A weeks is of course also highly correlated with the probability of having participated during, say, the last $z_A + 1$ weeks and in practice it is impossible to disentangle the precise nature of a dynamic effect pattern from the distribution of participation probabilities. Hence, the coefficient attached to A_{sjd} also needs to be interpreted with some care. Note that the variation in both N_{sjd} and A_{sjd} is basically county-by-duration-specific; the only reason why we include an s-subscript is that they vary slightly across individuals due to the leave-out-mean strategy described in (2); namely, that the intensities assumed to affect spell s are computed on the basis of all other spells in the same county and at the same duration.

To choose appropriate values of z_N and z_A, we estimate the hazard rate model under a number of alternative z_N-z_A-combinations and chose the combination that maximises the likelihood function. In order to save computational resources, this initial model competition is carried out without incorporating unobserved heterogeneity into the likelihood function (the inclusion of unobserved heterogeneity does in practice not appear to influence the ranking of models). The chosen model is then re-estimated

with full non-parametric specification of unobserved heterogeneity based on the strategy described in Appendix A. The results from our model competition exercise for the model based on full-time absence spells, with partial or full work resumption defined as the outcome variable are presented in Figure 3. With respect to the number of leads z_N on the notification variable, a clear 'winner' emerges, namely $z_N = 3$. This gives the highest likelihood regardless of how the attendance variable is specified. As $z_N = 3$ corresponds to our prior knowledge regarding the timing of the DM call (the call is indeed supposed to be sent out three weeks prior to the meeting), we find this reassuring. With respect to the duration of the attendance effect z_A, the results are less clear-cut (which is no surprise given the extremely high correlation we expect to find between measures based on small variations in z_A). Yet, for all choices of z_N, the likelihood is maximised for $z_A = 12$. Hence, our preferred model is based on the assumption that the notification of meetings occurs three weeks prior to their planed implementation and that the attendance effect lasts for 12 weeks. It should be noted, however, that our results would have been very similar had we selected other specifications in the neighbourhood of the chosen one.[7]

The main results based on the chosen model are presented in Table 2. Columns (1) and (2) show the results for the models with partial or full work resumption as outcome
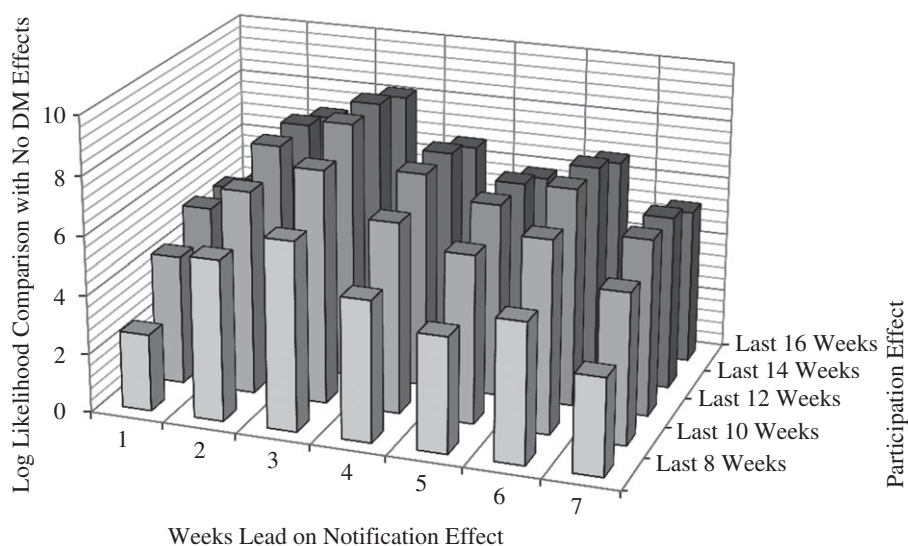

Fig. 3. *Choice of Model*

Notes. Sample of full-time spells with partial or full work resumption as outcome variable. The height of the columns represent the log likelihood for each of the 35 estimated models when subtracting the log likelihood from a model where both DM intensities A_{sjd} and N_{sjd} are excluded.

---

[7] The results of the model competition are very similar for the model estimated on all spells and full work resumption defined as the outcome, although it is in this case more difficult to identify a clear 'winner'. For this model $z_N = 2$ is marginally better than $z_N = 3$, and $z_A = 12$, $z_A = 14$ and $z_A = 16$ give almost exactly the same likelihood. To make the results directly comparable, we use $z_N = 3$ and $z_A = 12$ for both models/datasets.

Table 2

Results

	Partial of full work resumption		Full work resumption	
	(1) DM period	(2) Pre-DM period (placebo)	(3) DM period	(4) Pre-DM period (placebo)
Notification	1.691***	0.360	1.216***	0.230
	(0.600)	(0.658)	(0.447)	(0.537)
Attendance last 12 weeks	0.328***	0.174	0.169**	0.004
	(0.112)	(0.125)	(0.080)	(0.100)
No. of masspoints	12	10	10	11
Number of spells	2,014,576	1,880,895	2,115,424	1,966,940

Notes. *0.1 **0.05, ***0.01. Standard errors are given in parentheses. No. of masspoints refers to the number of masspoints in the distribution of unobserved heterogeneity. For a list of included control variables, see Appendix B.

variable for the pre-DM and DM periods respectively. Similarly, Columns (3) and (4) show the results for the models with only full work resumption as the outcome variable. Estimates of the same models for the pre-DM period may be interpreted as a sort of placebo analysis. The placebo regressions can only detect potential county confounders that are present both in the DM and the placebo time periods. They do, however, have the advantage of potentially revealing local treatment strategies (other than DMs) that targets sick-listed workers at particular durations of the spell that coincide with the typical timing of DMs. Given the ambiguities regarding the precise interpretation of the coefficient estimates, we use them primarily for statistical inference regarding the existence and sign of notification and attendance effects. Our discussion of magnitudes and substantive significance is instead based on simulation exercises where we compare results of imposing the meeting intensity profiles actually observed in different counties. Our results indicate statistically significant positive effects of both notification and attendance. The effects are considerably larger when the outcome is measured as an indicator for any degree of work resumption – partial or full – rather than an indicator for full work resumption only. Estimating on the sample of pre-DM spells instead (placebo analysis) gives much smaller and (in all cases) statistically insignificant coefficient estimates, supporting the causal interpretation of the estimates obtained based on spells from the period with DMs.

To shed some light on the magnitudes of the estimated effects, we perform a simulation exercise based on the estimated models where we simulate the duration of all absence spells under alternative assumptions regarding the two key variables N_{sjd} and A_{sjd} and their effects on the transition rates. More specifically, to ensure that we do not extrapolate outside the range of variation actually observed in the data, we compare simulated work resumption outcomes when we impose three alternative sets of DM profiles on N_{sjd} and A_{sjd}:

(*i*) Mixed DM intensity: the profiles actually estimated for each county.
(*ii*) High DM intensity: the profiles estimated for the county that uses DMs the most (Telemark).

    (*iii*) Low DM intensity: the profiles estimated for the county that uses DMs the least (Østfold).

To disentangle the quantitative impacts of notification and attendance, we repeat each simulation first with only the estimated notification effect and then with both of the estimated effects 'turned on' at the same time. The impact of attendance is then calculated from subtracting the impact of notification from the overall DM impact.[8] The results are presented in Table 3. The upper [lower] part of the Table shows results from the model, including full-time [all] spells. Columns (1) and (2) report summary statistics, while column (3) reports the effect on spell duration per meeting held implied by the simulation of the models. Looking first at the results for the model with

Table 3

Simulated Impacts on Duration Until Work Resumption of Alternative DM Profiles

	(1) Average spell duration (days)	(2) Fraction attending a DM (%)	(3) Implied effect on spell duration per extra meeting (days)
Partial or full work resumption (N = 2,014,576)			
Assumed DM profile			
Mixed (as estimated for each county)	29.16	1.35	
High (as estimated for Telemark)	29.04	1.93	
Low (as estimated for Østfold)	29.28	0.69	
Substituting high for mixed DM intensity	−0.12	+0.58pp	−20.3
Contribution from notification			−9.1
Contribution from attendance			−11.2
Substituting high for low DM intensity	−0.24	+1.24pp	−19.0
Contribution from notification			−8.4
Contribution from attendance			−10.6
Full work resumption (N = 2,115,424)			
Assumed DM profile			
Mixed (as estimated for each county)	39.18	2.42	
High (as estimated for Telemark)	39.07	3.47	
Low (as estimated for Østfold)	39.29	1.22	
Substituting high for mixed DM intensity	−0.11	+1.05pp	−10.8
Contribution from notification			−5.6
Contribution from attendance			−5.2
Substituting high for low DM intensity	−0.22	+2.25	−10.1
Contribution from notification			−5.3
Contribution from attendance			−4.8

Notes. The upper (lower) part of the Table shows results from the model, including full-time (all) spells. Columns (1) and (2) show summary statistics when imposing the three alternative assumptions on the DM profile, while column (3) is the effect on spell duration per meeting held implied by the simulation of the models.

[8] The impact of attendance is deduced in this way, rather than from a simulation with only the estimated attendance effect 'turned on', because a simulation with the estimated notification effect turned off gives too high a number of absentees actually attending a meeting and hence leads to an overestimation of the impact of attendance.

© 2017 Royal Economic Society.

partial or full work resumption defined as the outcome of interest, we note that the different DM profiles have very small impacts on overall absence duration. For example, substituting the high for the low DM profile reduces the average spell duration by a mere 0.24 days. However, this 'small' effect primarily reflects the fact that the DM profiles are relevant only to the tiny fraction of spells that lasts at least 5–6 months. Even the high DM intensity profile implies that less than 2% of the absentees ever attend a DM. Looking at the impact on average spell duration per realised meeting instead, we find that each extra DM reduces the duration until partial or full work resumption by as much as 19–20 days. Hence, for the target population of long-term absentees, the DM profile appears to be of considerable importance. Repeating the simulation exercise with only the notification effect 'turned on' and comparing the results, we find that the notification and attendance effects are of similar importance for average duration. The attendance effect is slightly larger than the notification effect. Moving on to the model with full work resumption as the outcome of interest, the simulated impacts are considerably smaller, yet far from negligible. Each extra DM is estimated to reduce the duration until full work resumption by around 10–11 days. And for this outcome, the notification effect is slightly larger than the attendance effect.

2.3. *Heterogeneous Effects*

By estimating one common effect for the entire sample, we might overlook heterogeneity across the distribution of sick-listed workers that could have important policy implications. To examine this, we divide the sample along a number of dimensions and estimate the model separately for each sub-sample. The dimensions related to the workers are gender, age and level of education.[9] A dimension related to the work place is firm size. Finally, we estimate heterogeneous effects according to whether or not the sick-leave certifier is also the family doctor. The results from the heterogeneity analyses are presented in Table 4. The correlations between the different dimensions by which we split the sample are ignorable, which implies a straightforward interpretation of the results. Columns (1)–(3) show the results for the model with partial of full work resumption as outcome variable and columns (4)–(6) show the results for the model with full work resumption as outcome variable. Columns (3) and (6) are *Z*-score for a test of equality of the estimated coefficients reported in columns (1) and (2), and columns (4) and (5) respectively. Overall, we find some, but quite modest heterogeneity of both notification and attendance effects in both models. The notification effect is considerably stronger for women than for men and we find indications of a stronger attendance effect for men. This might be related to differences in types of jobs across genders. Next, there are small or no significant differences across age groups and levels of education and no significant heterogeneity related to firm size. There is however, some heterogeneity related to whether or not it is the family physician who certifies the sick leave. In particular, the notification effect is stronger among those whose sick leave is certified by another physician than their

---

[9] The different dimensions by which we split the sample are defined in detail in the following notes below: Table 4.

Table 4

Heterogeneous Effects

	Partial of full work resumption			Full work resumption		
	(1)	(2)	(3)	(4)	(5)	(6)
(*a*)	Men	Women	Z-score	Men	Women	Z-score
Notification	0.314	2.675***	1.94	0.343	1.761***	1.53
	(0.933)	(0.785)		(0.735)	(0.565)	
Attendance	0.628***	0.127	2.19	0.204	0.163	0.25
last 12 weeks	(0.175)	(0.147)		(0.132)	(0.100)	
No. of masspoints	10	12		10	10	
Number of spells	859,707	1,154,872		886,080	1,229,347	
(*b*)	Young	Old	Z-score	Young	Old	Z-score
Notification	1.027	3.214***	1.69	1.115**	1.511*	0.41
	(0.0727)	(1.068)		(0.537)	(0.810)	
Attendance	0.349***	0.304	0.19	0.216**	0.059	0.90
last 12 weeks	(0.137)	(0.197)		(0.096)	(0.144)	
No. of masspoints	13	10		12	10	
Number of spells	1,496,335	518,244		1,565,764	549,663	
(*c*)	Higher edu.	Lower edu.	Z-score	Higher edu.	Lower edu	Z-score
Notification	2.025*	1.633***	0.02	1.273	1.253***	0.28
	(1.232)	(0.689)		(0.829)	(0.533)	
Attendance	0.760***	0.242*	0.31	0.140	0.195**	1.96
last 12 weeks	(0.230)	(0.130)		(0.148)	(0.095)	
No. of masspoints	10	12		10	12	
Number of spells	586,853	1,427,726		635,748	1,479,679	
(*d*)	Big firm	Small firm	Z-score	Big firm	Small firm	Z-score
Notification	1.796***	1.424	0.30	1.166**	1.293*	0.14
	(0.772)	(0.957)		(0.576)	(0.711)	
Attendance	0.291**	0.355**	0.27	0.078	0.317***	1.45
last 12 weeks	(0.145)	(0.179)		(0.102)	(0.129)	
No. of masspoints	14	9		11	9	
Number of spells	1,348,866	665,713		1,412,260	703,167	
(*e*)	Family physician	Other physician	Z-score	Family physician	Other physician	Z-score
Notification	1.278*	2.969***	1.31	0.851	2.371***	1.57
	(0.733)	(1.062)		(0.540)	(0.809)	
Attendance	0.349***	0.444**	0.39	0.186*	0.252*	0.37
last 12 weeks	(0.137)	(0.205)		(0.100)	(0.151)	
No. of masspoints	12	11		12	12	
Number of spells	1,270,756	743,823		1,345,479	769,948	

Notes. * 0.1 ** 0.05, *** 0.01. Standard errors in parenthesis. The main samples for the models with partial or full work resumption as outcome variables are divided into subgroups according to columns (1)–(2)/(3) for the model with partial or full work resumption as outcome variable and columns (4)–(5)/(6) for the model with full work resumption as outcome variable. The samples are split according to the following criteria: (*a*) The gender of the absentee. (*b*) Whether the absentee is younger than or older than age 50. (*c*) Whether the absentee has high education, defined as having a bachelor, masters or PhD degree. (*d*) Whether the absentee works in a firm with more than 30 employees. (*e*) Whether the family doctor or another physician certifies the sick leave. No. of masspoints refers to the number of masspoints in the distribution of unobserved heterogeneity. For a list of included control variables, see Appendix B.

family physician. This will typically be a physician from an emergency centre or a specialist form a hospital or polyclinic and their role in following up the patient might be unclear. The call to the meeting however, is usually sent to the family physician. This might explain why we observe a stronger DM effect when the certifier is another physician, since the DM call in these cases introduces some responsibility for the family physician in the follow-up work. Our results suggest that it is important to clarify the responsibilities of family physicians and specialists in the follow-up of the sick-listed workers.

2.4. *Assessment of the Causal Interpretation*

How sure can we be that the estimates reported in the previous subsection really represent the causal impacts of DM notification and attendance? What if counties that differ in terms of DM strategy also differ in the composition of absentees, in labour market conditions, or in their usage of other policy instruments? A first point to note here is that since we have included county-fixed effects in all our models, our identification strategy does not require county differences in absentee composition, local conditions, or in other policy instruments to be uncorrelated to their DM strategies. What it does require is that these presumed county differences do not exhibit the same duration pattern as the DM effects. To assess this potential challenge to our identification strategy, we first examine how the use of DMs correlates with potential confounders, such as medical diagnoses, waiting times for specialist treatment, alternative social insurance office initiatives and the local unemployment rate. As pointed out above, these confounders represent a potential threat to our identification strategy only to the extent that they influence the DM exemption decisions and/or have effects that coincide with the timing of DMs. In our examination of correlation patterns, we therefore focus on absence spells that potentially are exposed to DMs, namely, spells lasting for at least 20 weeks. We first check whether the county differences in DM propensity can somehow be explained by differences in absentee composition. The white bars of Figure 4 show the distribution of actual DM participation across counties for all absence spells lasting at least 20 weeks. The black bars show the corresponding distribution of DM participation conditional on a wide range of absentee characteristics, including 988 diagnosis dummy variables. As can be seen from this Figure, these two distributions are almost identical. Hence, the composition of absentees in terms of individual characteristics and medical diagnoses have no impact whatsoever on the distribution of DM intensities across counties. This confirms that the county differences in DM intensities do not stem from differences in absentee composition.

Counties that spend fewer resources on DMs may have more resources available for other purposes. A particular concern could be that moneys not spent on DMs are channelled to other programmes targeted at long-term sick-listed workers. If these competing programmes have negative effects on work resumption rates (e.g. due to lock-in effects), this could be the real story behind our (then misinterpreted) finding of favourable DM effects. In practice, the only alternative use of resources targeted at the same group of long-term absentees is to intensify the use of supported vocational rehabilitation (VR) programmes. These programmes normally do not start until
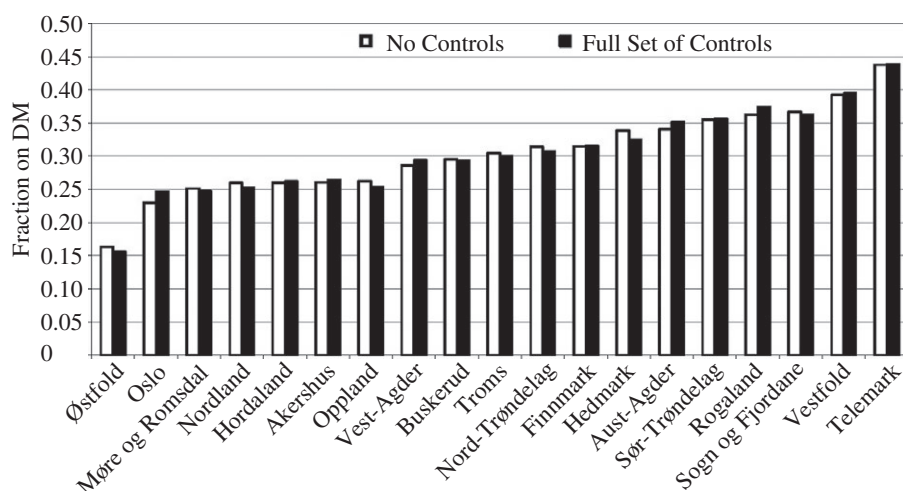
Fig. 4. *Conditional and Unconditional Distributions of DM Participation*
Notes. The white bars show the fractions of spells with a duration of at least 20 weeks where the absentee attends a DM. The black bars show the fractions after controlling for diagnosis and other covariates at the individual level through a linear regression. The regression includes 988 diagnoses dummy variables in addition to the individual explanatory variables described in Appendix B.

sick-pay entitlements are fully exhausted (i.e. after one year of absence) but they are sometimes started earlier. Hence, we cannot *a priori* rule out that there is a trade-off between DM intensity and rehabilitation programme intensity. To assess this concern, panel (*a*) of Figure 5 plots the county-specific DM rates for absence spells lasting at least 20 weeks (repeated from Figure 4) together with the corresponding rates of rehabilitation programme participation. There are two points to note from this Figure. The first is that very few long-term absentees (less than 2%) are enrolled in rehabilitation programmes; hence these programmes are unlikely to be relevant to the identification of the DM effects. And the second is that there appears to be no correlation between DM intensity and the use of rehabilitation programmes.

Another concern could be if the decisions to exempt absentees from DMs are motivated by the fact that they are waiting for some medical specialist treatment at the typical time of summoning. In that case, county variation in waiting times could lie behind some of the observed variation in DM intensity. To investigate this, we have collected data on the fractions of absentees who are waiting for specialist treatment at the typical time at which summoning decisions must be made (i.e. at 20 weeks absence duration) for each county. Panel (*b*) of Figure 5 again shows the county-specific DM rates for absence spells lasting at least 20 weeks, this time together with the county-specific fractions of the spells where the absentee is waiting for specialist treatment. As can be seen from this Figure, there is no indication that the county variation in DM intensity is explained by the county variation in the likelihood of waiting for specialist treatment. Finally, a third concern is the local SSA office has fewer resources available to assist absentees to return to work in periods of local recessions, and this affects the recovery profile of sick-listed workers particularly at the typical time of the meeting.
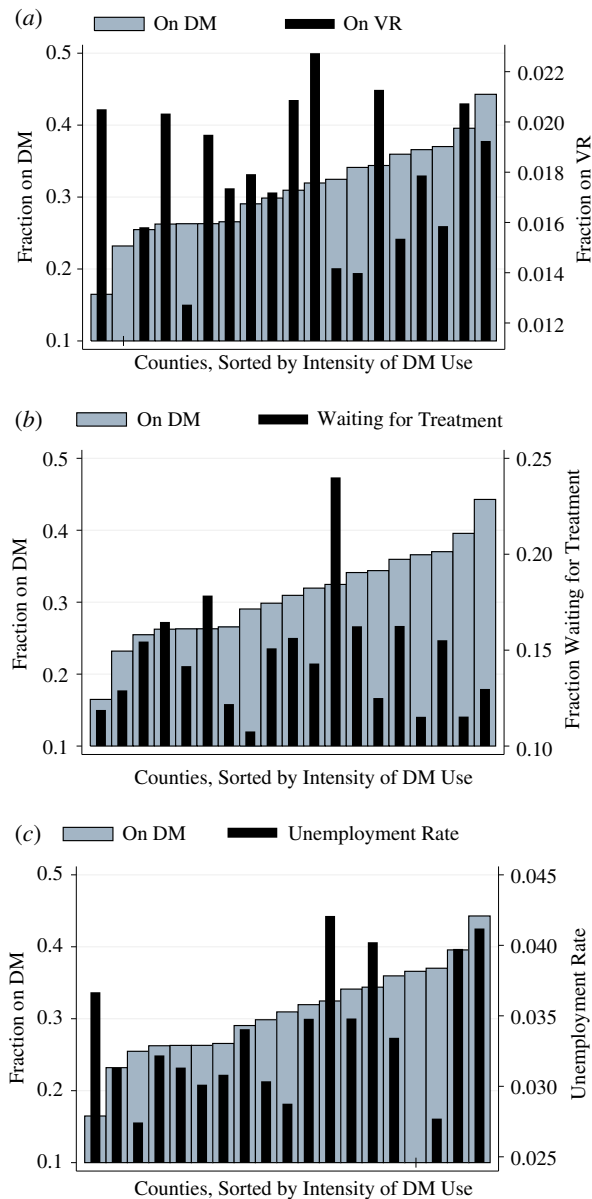
Fig. 5. *Potential Correlates of Counties' Use of DMs (a) DM use and Vocational Rehabilitation Programmes (b) DM use and Medical Treatment Waiting Times (c) DM use and Unemployment Rate*
Notes. Panel (*a*) shows the county-specific DM rates for absence spells lasting at least 20 weeks together with the corresponding county-specific rehabilitation programme participation rates. The correlation coefficient between these two county intensity measures is 0.13. Panel (*b*) shows the county-specific DM rates for absence spells lasting at least 20 weeks together with the corresponding county-specific fraction of spells where the absentee is waiting for specialist treatment. The correlation coefficient between these two county intensity measures is 0.01. Panel (*c*) shows the county-specific DM rates for absence spells lasting at least 20 weeks together with the corresponding county-specific unemployment rates. The correlation coefficient between these two county intensity measures is 0.28. Colour figure can be viewed at wileyonlinelibrary.com.

Table 5

Results with Flexible Controls for Local Unemployment Rates, Use of Vocational Rehabilitation Programmes and Medical Waiting Times

	Partial of full work resumption		Full work resumption	
	(1) Main results	(2) With additional controls	(3) Main results	(4) With additional controls
Notification	1.691***	1.687***	1.216***	1.318***
	(0.600)	(0.615)	(0.447)	(0.460)
Attendance last 12 weeks	0.328***	0.325***	0.169***	0.158***
	(0.112)	(0.114)	(0.080)	(0.081)
No. of masspoints	12	15	10	12
Number of spells	2,014,576	2,014,576	2,115,424	2,115,424

Notes. * 0.1 ** 0.05, *** 0.01. Standard errors in parenthesis. No. of masspoints refers to the number of masspoints in the distribution of unobserved heterogeneity. For a list of included control variables, see Appendix B.

Panel (*c*) of Figure 5 shows the county-specific DM rates for spells lasting at least 20 weeks together with county-specific unemployment rates. It does not seem to be any correlation between DM intensity rates and the local unemployment rate.

As an additional check to rule out any spurious correlations between the DM intensity measure and waiting times for specialist treatment, alternative social insurance office initiatives or the local unemployment rate, we include in the model duration specific controls for these potential county confounders. This is operationalised by creating ten indicator variables for each of the following spell duration intervals: week 1–4, 5–8, 9–13, 14–18, 19–22, 23–25, 27–32, 33–39, 40–49, 50–. These indicator variables are interacted with the county rate of rehabilitation programme participation, the county rate of waiting time for specialist treatment and county monthly measures of the local unemployment rate at the commuting zone level. The inclusion of these interactions controls for, in a highly flexible way, potential correlations between the local DM intensity and return-to-work opportunities at different durations of the spell, that result from changes in the use of vocational rehabilitation programmes, waiting times for specialist treatment or the local unemployment rate. The results from this exercise is presented in Table 5. It turns out that the inclusion of these controls does not affect our estimates to any noticeable extent.

3. Cost Efficiency

In order to assess cost efficiency, we compare the estimated DM gains with the associated costs. To obtain a measure of the benefits of the DM, we multiply the number of sick-leave days reduced per meeting by the average daily salary from the data set on all sick-leave spells (presented in Table 1). The average daily salary of the sick-listed workers in 2009 and 2010 is calculated to 1,580 NOK (US\$ 269). Our most conservative estimate of the number of saved sick-leave days is based on the model with only full work resumption counted as exit from sick-leave, and in this model we found

that each realised DM reduces sick-leave duration by approximately 10 days. Taking into account that this number also includes weekends, we estimate that each DM saves at least seven working days, or $1{,}580 \times 7 = 11{,}060$ NOK (US\$1,885). Since the replacement rate for most of the workers is 100%, the saved social insurance payouts will be of similar magnitude. Our cost estimates are based on information from the Norwegian social security administration (SSA). The SSA caseworker spends around three hours preparing and organising a meeting. With an annual employee cost of 650,000 NOK and 1,700 work hours per year, this corresponds to 1,150 NOK per meeting. The physician is on average compensated with 1,800 NOK per meeting. The employer is not compensated for DM attendance; however, we assume a cost of 1,800 NOK for the time of the employer. In total, this gives a cost of 4,750 NOK (US\$809). According to these crude calculations, the estimate of the benefits is more than twice as large as the estimate of the costs of arranging a meeting. And, notably, on the benefit side, this cost–benefit analysis includes only the added full-time work days. Based on the model for partial or full work resumption, we estimated that an extra seven working days can be added if we also count partial work resumption (since the most typical grade for partial absence spells is 50%, this constitute approximately 3–4 full-time days). Taken together, our results clearly substantiate that arranging compulsory DMs is a cost efficient way to reduce long-term absence.

4. Conclusion

This article examines the effect of a policy aimed at inducing long-term sick-listed workers in Norway to full or partial work resumption. The policy is a DM organised by the local social security administration around six months into sick-leave spells. To identify the causal effects of the DM, we exploit the fact that there is considerable geographical variation in both the overall use of DMs and in the timing of DMs within absence spells. We argue that this gives rise to random assignment-like variation in the duration-specific probabilities of being called to a DM, which makes it possible not only to identify the overall effects of the meeting but also to distinguish notification from attendance effects. In total, our results imply that each realised DM yields a reduction of around ten days in time until full work resumption, and a reduction of around 20 days in time until any form of work resumption (partial or full). The notification and attendance effects are of similar magnitude. Based on some simple cost–benefit calculations, we show that economic gains derived from earlier work resumption by far exceed the costs of arranging the meetings.

In our empirical analysis, we are not able to shed light on the relative importance of DM effects operating through the behaviours of the absentee, the physician and the employer. The meeting may be viewed as a monitoring device, whereby the social insurer can assess all three parties' efforts – or lack thereof – to ensure a successful return to work. The absentee may realise that the health situation no longer meets the requirements for continued full-time sick pay and thus take steps to resume work before this is established. The physician may respond to a call to a DM by not renewing a poorly justified absence certificate. Also the employer might respond by reconsidering the opportunities for modified work, as required by the law, even when it entails some additional costs. Although we cannot distinguish the contribution of the

absentee from that of the physician and employer, our conclusion is nevertheless the same; arranging mandatory DMs some time into an absence spell is an efficient policy measure to reduce long-term sickness absence.

The findings reported in this article may justify a more intensive use of DMs in Norway (at least up to the levels observed in the counties with the most intensive use so far) and possibly to introduce similar arrangements elsewhere. One question that naturally arises in this context is whether the DMs should be organised earlier than six months into the sick-pay spells. We obviously do not have any sound data-based foundation for evaluating such a policy. However, provided that the effects of DMs on work resumption hazards would have been the same for meetings held somewhat earlier, a simple simulation exercise indicate that arranging the meetings eight weeks earlier would have increased the number of meetings actually held by around 30–35% but increased the effect per realised meeting – in terms of reduced spell duration – by around 50%. It is possible, though, that a change in the timing of the meeting also changes its effects on the hazard rates. On the other hand, the exact timing of the meeting might not change its efficiency if the success of the meetings lies in their nudging of absentees and employers into breaking a pattern of procrastination of action.

Appendix A. Maximum Likelihood Estimation

The model of work resumption given in (6) is estimated by maximum likelihood. To do this in practice, each spell must be divided into a number of 'spell parts', such that each part is characterised by covariate constancy. This implies that any change in an explanatory variable (e.g. because a new duration week is beginning) triggers a new spell part. Let the subscript sp denote spell part p of spell s, let l_{sp} denote the length of that spell part (measured in days) and let y_{sp} be an indicator variable taking the value 1 if the spell part ended with a transition back to work and zero otherwise. The contribution of individual i to the likelihood function is then written:

$$L_i(\boldsymbol{v}_i) = \prod_{s \in S_i} \prod_{p \in s} [\theta_{sp,ijtd}(v_i)]^{y_{sp}} \exp\{-l_{sp}[\theta_{sp,ijtd}(v_i)]\}. \tag{A.1}$$

The vector of unobserved individual heterogeneity, $\boldsymbol{v}_i$ is approximated with an *a priori* unknown discrete probability distribution. The probability distribution is estimated non-parametrically by adding support points in the distribution until the model is saturated (Heckman and Singer, 1984). The preferred model is then chosen on the basis of the Akaike information criterion (AIC). See Gaure *et al.* (2007) for details on the estimation algorithm. Let Q be the number of support points in the distribution of unobserved heterogeneity v_m with associated probability q_m, $m = 1, 2, \ldots, Q$. The sample likelihood function can be written:

$$L = \prod_{i=1}^{N} \sum_{m=1}^{Q} q_m \prod_{s \in S_i} L_i(v_m), \quad \sum_{m=1}^{Q} q_m = 1. \tag{A.2}$$

Appendix B. Included Control Variables

The following conditioning variables are included in all models in Section 3.

   Spell duration (one category for each weekly duration; 53 categories).
   Calender time (one category for each calendar month; 36 categories).
   County (19 categories).
   Age (one dummy for each age from 18 to 66).

Native born (2 categories).

Diagnosis group. (In the main specifications: 31 categories based on average spell duration measured in five-days intervals up to a duration of 150 days, and then one category for diagnosis with an average spell durations exceeding 150 days. These are computed on the basis of sick-pay claimants who entered in 2008 and who are therefore not included in the analysis in this article.)

Level of education (10 categories).

Field of education (10 categories).

Income group (16 categories).

Sector/occupation (84 categories).[10]

Local unemployment rate (the share of the work force on unemployment benefits a given month, computed at the travel-to-work area level (40 regions in Norway); linear variable).

Appendix C. A Monte Carlo Assessment of the Identification Strategy Regarding DM Calls

Since we do not have data on meeting calls, an important strategy of this article has been to infer the county variation in meeting call intensities on the basis of the assumption that the call intensity is proportional to the subsequent realised meeting intensity. As pointed out in Footnote 6, this assumption is in general violated, as the ratio between the two intensities depends on the level of the involved baseline hazards, on the county-specific probability of participation, as well as on the size of the notification effect. Provided that there is a positive notification effect, the realised meeting propensity will also systematically underrate the call propensity. Using the realised meeting propensity as a proxy for the call propensity (as we do in this article) is therefore likely to induce a positive bias in estimated effect of each call. We have argued, however, that this error will have negligible effects on the simulation-based interpretation of the results. In this Appendix, we describe a Monte Carlo (MC) experiment that we have designed to assess the consequences of applying the county-specific realised meetings propensity as a proxy for the call propensity. The basic idea is to construct artificial data that have properties similar to those we use in the present article; and then check whether we are able to recover the 'true' notification and attendance effects based on the limited information available in the actual data set. Somewhat simplified, a baseline version of the MC experiment is designed as follows:[11]

(i) We start by constructing a vector of 53 ('weekly') probabilities ('recovery rates') equal to the average observed weekly recovery rates observed in the actual data. These probabilities are constructed directly from the estimation dataset.

(ii) We generate 250,000 unique identification numbers each corresponding to a 'spell'. All these spells are randomly allocated into three equally sized entities ('counties') with equal probabilities. To start with, these counties play no role and, since there are no meetings in the first weeks, we start the simulation in week 17.[12]

(iii) For each spell, we draw a number from a [0, 1] uniform distribution and compare it to the week 17 average recovery rate. If the number is smaller than the recovery rate, we

---

[10] There has been a change in the sector coding over the time period of study. This means that the sector codes differ for the data from the pre-DM and the DM periods. In the model on long-term outcomes, the sector coding is hence different for spells started in the two different time periods.

[11] The exact algorithm can be obtained from https://www.dropbox.com/s/vohttgo9uipfwm6/Monte, and the data set from https://www.dropbox.com/s/z98nskblnt3zd9m/hazsnitt.dta?dl=0

[12] We also ran the simulation starting in week 1. This gives the same conclusions but requires considerably more computational resources since a much larger sample is needed to ensure the same number of observations for which the meeting becomes relevant (with survival to the week of notification).

record this number as corresponding to a spell of 17 weeks duration. If the spell does not end, we repeat the same exercise for week 18 and (with the week 18 and 19 average transition rates, respectively).

(*iv*) For 'survivors' who have still not made a 'recovery' after week 19, we randomly allocate the spells to DM calls. The DM call probabilities vary between the three counties with the following DM call probabilities: 0.18, 0.36, 0.50, corresponding roughly to the DM propensies observed in the actual data for the two most extreme counties and the national average. If a DM call is drawn, the recovery rate in week 20 is raised by a factor of $\exp(1.50) = 4.5$ (the notification effect).

(*v*) We continue drawing recoveries in week 21 and 22, based on the original vector of recovery rates.

(*vi*) Then, for those who were drawn to receive a DM call in week 20, but have not yet 'recovered' at the end of week 22, we multiply the elements of the recovery rate vector for weeks 23–34 by $\exp(0.33) = 1.4$ (the attendance effect), and continue drawing recoveries.

(*vii*) From week 35, we again return to the original vector of recovery rates until week 53, after which the still ongoing spells are right-censored.

These seven steps constitute our data generating process (DGP). We then use the hazard rate strategy described in Section 3 to estimate the impacts of the DM, just as we do with the actual data. That is, we pretend that we do not observe the meeting calls in week 20 and use instead the leaded county-specific leave-out means for realised meetings in week 23 to proxy the county-specific call probability three weeks earlier. To proxy the county-specific probabilities of already having participated, we use the observed leave-out means. The results from this exercise are presented in Table C1. We report both the estimated coefficients in the hazard rates, and the corresponding simulated effects on spell duration (in days). To illustrate the statistical uncertainty involved, we have repeated the whole DGP and estimation process 1,000 times, and report in the table the mean results/estimates together with the standard deviations over the

Table C1

Estimation and Simulations Results Based on Generated Data in 1000 Trials, Each with 250,000 Spells

	(1) True effect in DGP	(2) Estimated effect based on data with correct individual information on calls and realised meetings	(3) Estimated effect based on data with information on realised meetings only and with county-based leave out means to predict individual attendance
Notification coefficient in hazard rate	1.50	1.50 (0.03)	1.69 (0.11)
Attendance coefficient in hazard rate	0.33	0.33 (0.01)	0.34 (0.06)
Total effect on average duration per realised meeting (days)	−39.3 (3.80)	−39.3 (4.61)	−39.4 (4.57)

Notes. Standard deviations in parentheses. The true effects and simulation results based on the true parameters in the DGP and the implied reduction in absence per realised meeting are reported in column (1). Column (2) reports the results obtained when using the correct individual information on meeting calls. Column (3) reports the estimated effects based on data with information on realised meetings only which are based on the estimation strategy used in this article; i.e. with county-specific leave out means used to proxy for DM attendance (with actual attendance rates in week 23 used as proxy for the call propensity in week 20).

1,000 trials (in parentheses). The true parameters in the DGP and the implied reduction in absence per realised meeting are reported in column (1). In column (3) we report the results based on the estimation strategy used in the present article, where call intensities are not observed. For comparison, we also report the results obtained when using the correct individual information on meeting calls to estimate the effects within the same type of hazard rate model (column (2)).

As expected, our empirical strategy systematically overestimates the notification effect in the hazard rate (1.69 instead of 1.50) but estimates the attendance effect correctly. The simulation-based duration-impacts of notification and attendance appear to be consistently estimated. It is notable that our empirical approach based on county-specific leave-out-means of actual attendance propensities are as consistent as the model where we observe not only on attendance, but also calls. The only difference of importance is that the standard deviations of the estimated coefficients are much larger for the model without this information. This has little impact on the simulation results, however. Simulations based on the model used in this article are actually equally informative as simulations based on the data where we observe also meeting calls. Our empirical strategy of using the realised meeting propensity as a proxy for the call propensity three weeks earlier requires that the hazard rates are not defective. If, for example, the notification is infinitely large, such that everyone receiving a call returns to work immediately, we will never observe any realised meetings and hence have no information about the call intensity. However, it turns out that our strategy works well even when the notification effect is extremely large, such that almost everyone receiving a call returns to work immediately. For example, when we repeat the whole exercise described above with a notification coefficient equal to 4 (such that the recovery hazard is shifted upwards by a factor of $\exp(4) = 55$), we find that our estimation strategy grossly exaggerates the notification effect (12.5 instead of 4.0). Yet it still yields a correct attendance effect. And, importantly, it still gives a correct simulation-based estimate on the overall effect per meeting. Since the attendance effect is so large in this experiment, very few spells ever reach the time of an actual meeting; hence in the DGP, the average duration effect per realised meeting is as large as 1,212 days. Based on our empirical strategy, we would have simulated an effect of 1,224 days.

Ragnar Frisch Centre for Economic Research

References

Black, D.A., Smith, J.A., Berger, M.C. and Noel, B.J. (2003). 'Is the threat of reemployment services more effective than the services themselves? Evidence from random assignment in the UI system', *American Economic Review*, vol. 93 (4), pp. 1313–27.

Cockx, B., Ghirelli, C. and Van der Linden, B. (2014). 'Is it socially efficient to impose job search requirements on unemployed benefit claimants with hyperbolic preferences?', *Journal of Public Economics*, vol. 113, pp. 80–95.

DellaVigna, S. and Paserman, M.D. (2005). 'Job search and impatience', *Journal of Labor Economics*, vol. 23(3), pp. 527–88.

Dolton, P. and O'Neill, D. (1996). 'Unemployment duration and the restart effect: some experimental evidence', Economic Journal, vol. 106(435), pp. 387–400.

Fevang, E., Markussen, S. and Røed, K. (2014). 'The sick pay trap', *Journal of Labor Economics*, vol. 32(2), pp. 305–36.

Gaure, S., Roed, K. and Zhang, T. (2007). 'Time and causality: a Monte Carlo assessment of the timing-of-events approach', *Journal of Econometrics*, vol. 141(2), pp. 1159–95.

Geerdsen, L.P. (2006). 'Is there a threat effect of labour market programmes? A study of ALMP in the Danish UI System', Economic Journal, vol. 116(513), pp. 738–50.

Gorter, C. and Kalb, G.R. (1996). 'Estimating the effect of counseling and monitoring the unemployed using a job search model', *Journal of Human Resources*, vol. 31(3), pp. 590–610.

Graversen, B. and Larsen, B. (2013). 'Is there a threat effect of mandatory activation programmes for the long-term unemployed?', *Empirical Economics*, vol. 44(2), pp. 1031–51.

Hägglund, P. (2011). 'Are there pre-programme effects of active placement efforts? Evidence from a social experiment', *Economics Letters*, vol. 112(1), pp. 91–3.

Heckman, J. and Singer, B. (1984). 'A method for minimizing the impact of distributional assumptions in econometric models for duration data', *Econometrica*, vol. 52(2), pp. 271–320.

Henrekson, M. and Persson, M. (2004). 'The effects on sick leave of changes in the sickness insurance system', *Journal of Labor Economics*, vol. 22(1), pp. 87–114.

Hägglund, P. (2011). 'Are there pre-programme effects of active placement efforts? Evidence from a social experiment', *Economics Letters*, vol. 112(1), pp. 91–3.

Johansson, P. and Lindahl, E. (2013). 'Can sickness absence be affected by information meetings? Evidence from a social experiment', *Empirical Economics*, vol. 44(3), pp. 1673–95.

Johansson, P. and Palme, M. (1996). 'Do economic incentives affect work absence? Empirical evidence using Swedish micro data', *Journal of Public Economics*, vol. 59(2), pp. 195–218.

Markussen, S., Mykletun, A. and Røed, K. (2012). 'The case for presenteeism – Evidence from Norway's sickness insurance program', *Journal of Public Economics*, vol. 96(11–12), pp. 959–72.

Markussen, S. and Røed, K. (2014). 'The impacts of vocational rehabilitation', *Labour Economics*, vol. 31(0), pp. 1–13.

Markussen, S. and Røed, K. (2016). 'The Market for Paid Sick Leave'. Discussion Paper No. 9825, IZA.

Markussen, S., Røed, K. and Røgeberg, O. (2013). 'The changing of the guards: can family doctors contain worker absenteeism?', *Journal of Health Economics*, vol. 32(6), pp. 1230–9.

OECD (2010). *Sickness, Disability and Work: Breaking the Barriers*, Paris: OECD.

Paserman, M.D. (2008). 'Job search and hyperbolic discounting: structural estimation and policy evaluation', ECONOMIC JOURNAL, vol. 118(531), pp. 1418–52.

Rosholm, M. and Svarer, M. (2008). 'The threat effect of active labour market programmes', *Scandinavian Journal of Economics*, vol. 110(2), pp. 385–401.

Van den Berg, G.J. (2001). 'Duration models: specification, identification and multiple durations', in (J. Heckman and E. Leamer, eds.), *Handbook of Econometrics*, pp. 3381–460, vol. 5 of Handbook of Econometrics, chap. 55, Elsevier, Amsterdam.

Vågeng, S., Eriksen, H.R., Ihme, I., Markussen, S., Pedersen, N., Stene, E. and Sæther, I.(2015). 'Et NAV med muligheter.', *Sluttrapport fra Ekspertgruppe*. https://www.regjeringen.no/globalassets/departementene/asd/dokumenter/2015/sluttrapport-ekspertgruppen-nav_9.4.15.pdf (last accessed :17 February 2017).