Activation against absenteeism - Evidence from a sickness insurance reform in Norway

Hernæs, Øystein M.

Journal of Health Economics 62 (2018) 60–68

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Activation against absenteeism – Evidence from a sickness insurance reform in Norway[☆]

Øystein Hernæs

The Ragnar Frisch Centre for Economic Research, Gaustadalleen 21, 0349 Oslo, Norway

ARTICLE INFO

Article history:
Received 21 December 2017
Received in revised form
19 September 2018
Accepted 20 September 2018
Available online 27 September 2018

JEL classifications:
I18
I38
J48

Keywords:
Absenteeism
Disability
Activation
Workfare

ABSTRACT

I evaluate a program aimed at strictly enforcing a requirement that people on long-term sick leave be partly back at work unless explicitly defined as an exception. Employing the synthetic control method, I find that the reform reduced work-hours lost due to sickness absenteeism by 12% in the reform region compared to a comparison unit created by a weighted average of similar regions. The effect is driven by both increased part-time presence of temporary disabled workers and accelerated recovery. Musculoskeletal disorders was the diagnosis group declining the most. The findings imply large savings in social security expenditures.

1. Introduction

Can an activation policy targeted at making use of the partial work capacity of temporary disabled workers reduce long-term absenteeism and bring down social security costs?

Disability rolls have been rising for decades in many OECD countries, entailing both a substantial volume of labor withdrawn from the market, as well as heavy social security costs (OECD, 2010). This has led to increased attention on the trade-off between generosity towards those hit by a negative health shock and potential moral hazard problems that face any social security system. Traditional responses to this trade-off have been to establish strong screening criteria or other gatekeeping policies, or to limit the level or duration of benefits. In this paper, I analyze the effectiveness of an activation strategy based on graded sickness insurance, i.e. requiring temporary disabled workers to be partly back at work to the extent possible, as opposed to full-time absence.

Since 2004, the Norwegian social insurance system has required employees on long-term sick leave to be partly back at work unless explicitly judged by a physician to be unable to work at all, irrespective of adaptions at the workplace. This policy encourages *presenteeism*, i.e. being present at work despite being sick. However, despite the clear letter of the law and pronounced political intentions, this rule has not been rigorously enforced. I analyze a program implemented in the Norwegian region of Hedmark in 2013 aimed at more strictly enforcing this already existing requirement.[1] There are clear indications that the reform was successful in reducing absenteeism: a report from the Labour and Welfare Administration computed a fall of 8% in the absence rate in Hedmark relative to the countrywide average in the year following the reform (Kann et al., 2014). In this paper, I conduct a rigorous analysis of a longer time period, consider the uncertainty and robustness of the results, and go into mechanisms.

The fact that the program was comprehensively implemented across a whole region in a top-down fashion makes it natural to base an effect evaluation on some sort of comparison with other regions. Thus, the choice of comparison group becomes crucial. To

☆ Funding from the Røwde foundation and the project "Explaining social patterns in sicknes absence: the influence of values, attitudes and norms" (Norwegian Research Council grant #237993) is gratefully acknowledged. I wish to thank Tarjei Havnes and Knut Røed for valuable comments at earlier stages of the project.

E-mail address: o.m.hernas@frisch.uio.no

[1] Norway has 19 regions with an average population of 269,000. Hedmark has 194,000 inhabitants.

https://doi.org/10.1016/j.jhealeco.2018.09.007
0167-6296/

make this choice, I employ the synthetic control method (SCM) due to Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2015), which provides a data-driven approach to the selection of a comparison group. The SCM matches the treatment region Hedmark and the control regions on several pre-treatment variables, in order to construct a counterfactual "synthetic" Hedmark based on a convex combination of regions. The "synthetic" Hedmark is similar to the actual Hedmark, except that the synthetic region did not implement any changes to activation policies.

The comparison of the actual region versus the constructed counterfactual shows that enforcing the graded sick-leave requirement reduced the sickness absence rate by 12%. Systematic placebo tests on all other regions and for an earlier time period find no estimated effects of comparable magnitude. The effect is also robust to successively leaving out the regions receiving a positive weight in the baseline analysis. I find evidence that the absence rate declined not only through exploiting the partial work capacity of temporary disabled workers, but also by speeding up the transition rate back to full-time work. Consistent with expectations, the largest decline occurred for absenteeism due to musculoskeletal disorders, the smallest for respiratory disorders, with diagnoses for psychological and other disorders in between. The findings imply savings in social security expenditures of USD 310 per employee per year. I also find substantial effects of a similar reform introduced later by another region, which shows the program to be generalizable.

The key finding of the paper is that making use of the partial work capacity of temporary disabled workers has the potential to reduce long-term absenteeism and bring down social security costs. The paper contributes to the literature by quantitatively demonstrating that this can be achieved by relatively simple means by social security administrations working within existing legislation and contractual obligations. This is also the first paper to investigate with population register data how different patient groups respond to such activation requirements. The beneficial results for most diagnosis groups are encouraging, in particular the large effect for musculoskeletal disorders, which is responsible for a large and growing share of sickness absenteeism.

2. Related literature

The paper is related to the literatures on activation strategies and return-to-work interventions for sick or temporary disabled workers, see Røed (2012), van Vilsteren et al. (2015) and Holm et al. (2017) for reviews. Graded sick leave has been growing in popularity as a policy tool in recent years, and its impact has been investigated in several Northern European countries. Høgelund et al. (2010, 2012) and Schneider et al. (2016) find that being on graded sick leave is associated with a substantially higher chance of returning to regular working hours in Denmark and Germany, respectively. Likewise, Andrén and Svensson (2012) and Andrén (2014) show positive associations between graded and return to full-time work also for a sample of Swedish sick-listed employees. Høgelund and Holm (2014) find a positive association between workplace adaptions and employment durations for sick-listed workers in Finland. Analyzing a small (N = 63), randomized controlled trial in Finland, Viikari-Juntura et al. (2012) find that grading accelerated the return to regular work of people suffering from musculoskeletal disorders. This intervention did not affect a range of specific health measures, but improved self-rated general health and health-related quality of life (Shiri et al., 2013). Also in Finland, Kausto et al. (2012) and Kausto et al. (2014) use propensity score matching to compare people on partial vs. full-time sick leave, and find that partial sick leave is associated with a larger share of people on partial rather than full-time disability pension and a higher degree of employment. The present study contributes to this literature by employing population-level data and relying on less strong modeling assumptions.

Evidence is mixed on how people on long-term absence respond to benefit cuts (Johansson and Palme, 2002, 2005; Ziebarth, 2013), however, activation strategies may present a promising policy tool also for this group. A contribution closely related to the present paper is Markussen et al. (2012), who also study activation requirements for workers on long-term absence in Norway. Using administrative register data from the first half of the 2000s, they employ an instrumental variable approach based on variation in physicians' propensity to use graded absence certificates in contrast to awarding full-time absence certificates. Their results are encouraging – requiring workers on long-term sick leave to work part-time decreases absenteeism and raises later employment propensities. The results are highly policy-relevant, as they are well identified and based on a scalable treatment that consists of a broad activation policy not restricted to certain therapies or diagnoses. This paper complements Markussen et al. (2012) in three important respects: First, they estimate a local average treatment effect based on patients whose grading outcomes are influenced by the physician, whereas in this paper, the estimate is the average treatment effect for everyone in the region. Second, although the results are stable across a range of robustness checks, the chance remains of positive selection of patients to physicians with higher grading propensities, or that unobserved physician factors might influence both grading strategies and patient outcomes. Third, their main analysis is conducted in a context where graded absence certificates were relatively rare and most absences spells were 100%. Indeed, when they restrict their sample to after July 2004, when a reform sharply increased the use of graded absence certificates, the estimated effects are smaller, raising the worry of diminishing returns to the favorable effects of grading. It is therefore reassuring that this paper, which is based on a completely different identification strategy, in a setting in which graded absence certificates were much more common, also finds substantial beneficial effects. A final innovation is to analyze how different patient groups, in particular differentiated by diagnoses, respond to the reform.

This paper is complementary to the literature on the costs of contagious presenteeism (Chatterji and Tilley, 2002; Barmby and Larguem, 2009; Pichler and Ziebarth, 2017) in that the policy it evaluates is explicitly targeted at workers with non-contagious diseases who retain some degree of work capacity.

3. Activation reform

In the Norwegian sickness insurance system, the replacement ratio is 100% from the first day of sick leave, with a maximum duration of one year.[2] The employer pays for the first 16 days, while the state covers the period thereafter. For absence spells of more than 3 days, a physician must certify in writing that the employee has a medical condition that prevents work. To some extent, the incentives of the parties involved are stacked against a return to work for a sick-listed employee. The financial risk associated with a relapse weakens the incentive of the employer to get the employee back to work. An absence spell is counted as a continuation of a previous spell if the employee has been back at work for less than 16 days. This reduces, though does not eliminate, the risk that the employer will become liable to cover a second period of absence. Fevang et al. (2014) investigated this disincentive effect and found it to have a substantial, negative effect on the transition back to work. A sick-listed employee with a salary less than the maximum amount

[2] The replacement level is 100% of the salary up to a yearly salary of 6G, where "G" is the basic amount in the National Insurance scheme and 1G = NOK 96 883 ≈ USD 11 600.

covered also does not have a short-term monetary incentive to go back to work. Finally, the fact that people are free to choose their own physician, who have an economic interest in retaining their patients, creates a "market for paid sick leave" encouraging leniency on part of the physician (Markussen and Røed, 2017). This is important, as the physicians are supposed to serve a gatekeeping role in the sickness insurance system.

As a measure to limit moral hazard problems on part of both the employer and the employee, as well as to accelerate recovery (Mykletun et al., 2010), the Norwegian sickness absence system has since 2004 included an activation requirement. This requirement stipulates that people on sick leave should be partly back at work as early as possible and at the latest after 8 weeks. This is based on the fact that most long-term sick-leave days are caused by musculoskeletal diseases, back pain, mental disorders or other illnesses that rarely are completely disabling and do not risk infecting other workers (Markussen et al., 2012), and for which some degree of activity may even be healthy (Hagen et al., 2003; Waddell and Burton, 2006). In contrast, short-term absence spells are primarily due to respiratory infections and virus and gastrointestinal diseases, for which grading is much less relevant (Markussen et al., 2011). The most recent analysis of the relation between graded share and spell duration shows that the share of absence spells that is graded indeed rises sharply with spell duration, from 7% at one week to 31% at 8 weeks, after which it eventually plateaus at around 40% at 40 weeks (Markussen et al., 2012).

The activation requirement is aimed at exploiting the remaining work capacity of the temporary disabled worker. When writing an absence certificate, which is necessary to receive sick-pay for absence spells of more than 3 days, physicians are obliged to report the fraction of the patient's work capacity that is lost due to the illness. A graded certificate indicates that this fraction is less than 100%. In the case of 30% lost work capacity, the worker is obliged to perform 70% of her/his regular work duties, for which regular wages are paid, and receive sick pay for the remaining 30%. Implementing this at the workplace typically involves a dialogue with the employer and some degree of workplace- and individual-specific adaptions.

Within 8 weeks of sick leave, the local social security administration is supposed to decide formally on whether a case is an exception to the activation requirement. Exceptions may be granted either based on medical grounds, or because it is impossible to implement suitable adaptions at the workplace in question. An exception due to medical reasons should be based on a written explanation by the employee's physician detailing why the patient's remaining work capacity is 0%. An exception due to the impossibility of implementing adaptions at the workplace should rest on an explanation by the employer for why workplace adaptions are not possible. In 96% of exceptions in 2012, medical reasons were given as the basis for the exception (Brage et al., 2014).

In the fall of 2012, the social security administration in the region of Hedmark undertook an extensive investigation into how sick-listed workers were followed up by their local social security office.[3] There is one such office in each municipality, of which there are 22 in Hedmark. The investigation found that the requirement that an absentee should be partly back at work by 8 weeks, to the extent possible, had been far from rigorously enforced: The social security administration often did not formally decide on exceptions at the 8-week point and accepted incomplete explanations for exceptions from physicians and employers. On the basis of this knowledge, the regional social security administration devised a reform designed to enforce the activation requirement comprehensively and consistently throughout the region, starting from the second quarter of 2013. The reform consisted only of changing the practice of existing laws, not of changing any laws themselves.

The most important practical element of the reform was the introduction of weekly meetings in all local offices where all sick-leave cases approaching the 8-week mark were to be discussed. In these meetings, caseworkers discussed their cases with administrative support staff from the regional office and a physician and a psychologist employed by the social security administration. The meetings ensured that the required documentation for exceptions had been provided, that possibilities for making use of the individual's a remaining work capacity had been explored and that all cases were formally decided on. The presence of a physician and a psychologist ensured that the decision was safe and reasonable from a medical point of view. Always requiring sufficient documentation likely helped limit moral hazard problems between the worker and the physician, as they now had to discuss and document concretely why no amount of work could be performed. Previously, in cases with some degree of remaining work capacity, there was more of an opportunity for both to collude to act on their private incentive for 100% absence, as discussed above, by not entering into an explicit justification for why no amount of work was possible. After the reform, such a form of collusion would require outright misrepresentation in the documentation.

There was also a change in the communication from the social security office, consisting of rewriting a letter supposed to be sent to (full-time) absentees approaching 8 weeks absence. The new letter included information about termination of the sick leave benefit unless satisfactory documentation for an exception were provided and encouraged contact with the employer rather than the physician going forward. After the reform, this letter was sent to all absentees whose cases lacked sufficient documentation. Previously, a letter was sent in only around 5% of cases. An overview of typical practice before and after the reform is provided in Table A.1 in the Appendix in Supplementary material.

The regional social security adminstration took several steps to prepare the parties involved, including meeting with all local offices and physicians and psychologists employed by the social security administration, sending letter and e-mail about the changes to all physicians in the region, using the region's "workplace support center" to inform employers, and informing the general population through local newspapers and other media.

4. Method and data

The activation reform was comprehensively implemented across the whole region of Hedmark from the second quarter of 2013. From a policy evaluation perspective, the fact that the implementation took place at a specific point in time determined by the central authorities is helpful, as it allows a credible comparison of the pre- and post-periods. Yet, absenteeism levels move substantially over time also for other reasons, necessitating a comparison of the development from pre- to post-period with developments elsewhere.

One way to do to make these comparisons would be a difference-in-differences (DiD) analysis, comparing the means of the treated and non-treated regions across time and space. When computing standard errors in panel data settings with regionally based treatments, it is important to account for the possibility of serial correlation brought about by regional shocks (Bertrand et al., 2004). The most common way of doing this is to cluster standard errors at the level of the treatment (Angrist and Pischke, 2009). In this application, that would mean to cluster by regions, of which there are 18 in the sample. This would be the case even if one had data on a lower geographical level, such as municipalities or individu-

[3] The background for and account of the reform is based on NAV Hedmark (2014).

als. However, these standard errors are only asymptotically valid and are biased in cases with a small number of clusters (Angrist and Pischke, 2009). The problem of inconsistent standard errors becomes worse when the treatment is concentrated in few clusters (Conley and Taber, 2011), leading Cameron and Miller (2015) to suggest the synthetic control method (SCM) with aggregate data as an alternative in such cases.

Further, a simple average of other regions may not constitute the most relevant comparison group, and though the DiD framework does the important job of differencing out individual-specific heterogeneity that is fixed in time, time-varying factors at the individual level may also be important.

To deal with these concerns, I employ the synthetic control method (SCM) due to Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2015) to construct the counterfactual. The goal is to estimate the effect of the intervention/treatment for unit i at time t: $\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$, where Y^{I} denotes the outcome of interest in case of exposure to the intervention, and Y^{N} denotes the outcome in the absence of intervention. The SCM is motivated by a structural model in which the non-intervention outcome for a unit i is given by:

$$Y_{it}^{N} = \delta_t + \boldsymbol{\theta_t Z_i} + \boldsymbol{\lambda_t \mu_i} + \varepsilon_{it} \tag{1}$$

where δ_t is a common time-dependent factor, $\boldsymbol{\theta_t}$ is a vector of unknown parameters, $\boldsymbol{Z_i}$ is a vector of observed covariates not affected by the intervention, $\boldsymbol{\lambda_t \mu_i}$ is a vector of unobserved time-specific common factors multiplied by a vector of unobserved unit-specific factor loadings, and ε_{it} is an unobserved transitory shock. Note that the term $\boldsymbol{\lambda_t \mu_i}$ allows the effect of unobserved unit-specific confounders to vary over time. Next define the unit of interest as unit $i = 1$, and a vector of weights $\boldsymbol{W} = (w_2, \ldots, w_{J+1})'$, such that $w_j \geq 0$ and $w_2 + \ldots + w_J = 1$, defined over all J non-treated units. Any $\mathbf{W}$ represents a possible synthetic control and can for any time period be used to produce a linear combination of non-treated units $\sum_{j=2}^{J+1} w_j Y_{jt}$ to serve as a counterfactual for the intervention unit. Abadie et al. (2010) show that in the context of the model specified in Eq. (1), a $\mathbf{W^*}$ such that $\sum_{j=2}^{J+1} w_j^* Z_j = Z_1$ and $\sum_{j=2}^{J+1} w_j^* \bar{Y}_j = \bar{Y}_1$, where $\bar{Y}_i$ is a linear combination of pre-intervention outcomes, would provide an unbiased estimator of Y_{1t}^{N}. This solution is approximated by minimizing the distance between these so-called predictor variables of the treated unit and the synthetic control unit with respect to $\mathbf{W^*}$. In this procedure, the weights put on the predictor variables are chosen to minimize the mean square predictor error of the outcome variable for the pre-intervention (or pre-validation) periods.

The result is a counterfactual consisting of a convex combination of non-treated ("donor") regions. In this paper, I employ averages of the outcome variable and the workforce and age structure of the region as the predictor variables to be used in the construction of $\mathbf{W^*}$. These averages are defined over the period ranging from the first quarter of 2008 until the fourth quarter of 2011. This is also the period for which the mean squared prediction error of the outcome is minimized. The four quarters of 2012 and the first quarter of 2013 serve as a validation period for the weights constructed from the earlier period, before the treatment is introduced in the second quarter of 2013.

The procedure described above provides an estimate of the counterfactual development of the treatment region, and thus of the treatment effect for that region, however, it does not say anything about the uncertainty of that estimate. As the number of regions is relatively small, standard large-sample inference is not applicable. I follow Abadie et al. (2010), who based on the idea of a permutation test suggest systematic placebo tests on all untreated units to see whether the estimated effect for the treatment unit is large relative to the placebo effects for the untreated units. Following Abadie et al. (2015), I complement these placebo tests "in space" with a placebo test "in time," moving the reform forward to the middle of the pre-treatment period, as well as a robustness exercise consisting of leaving out the regions receiving positive weight in the baseline results.

All variables are region-level data aggregated from individual-level register data by Statistics Norway. The main outcome is the sickness absence rate, defined as "person-days lost due to own sickness as a percentage of contractual person-days." This rate expresses the amount of work hours lost to physician-certified absenteeism based on measuring the "stock" of physician-certified sickness absence on specific reference days, taking into account contracted work hours and degree of absenteeism. This measure includes all absence spells of more than 3 days. Absences of 3 days or shorter are self-certified by the employee and not part of administrative registers. I use data starting in 2008, as there was a break in the data series at that point regarding industry classification scheme. However, the reform that I study was not implemented until the second quarter of 2013, which leaves 21 quarterly pre-treatment observations. I study the period up to and including the fourth quarter of 2015. In the main analysis I focus on all employees (i.e. everyone aged 16–69), but I also break the analysis down by gender, age, sector and diagnosis group. It should be noted that as the outcome or subgroup changes, the synthetic control will typically also change, thus these estimates should be interpreted with this in mind. To provide some information about the uncertainty also of these secondary estimates, the estimated effects will be displayed visually together with the placebo estimates for the other units.

The covariates on which I match are the share of inhabitants with a university or college education, the share of the workforce employed in the health/social sector and the public sector, the female share of the workforce and the age composition of the inhabitants.

There are 19 regions in Norway. The regions are the second administrative level of the country, below the national level. The main responsibilities of the region governments are upper secondary education, regional roads, local public transportation and some cultural and health services (Fiva and Halse, 2016). The Norwegian Labour and Welfare Administration has offices in all 428 municipalities of the country (the third administrative level), however these are internally organized under a region-level leadership, which may exert considerable influence on practice in the municipality offices. I am forced to exclude the region of Oppland, as it introduced a similar activation program some time later, thus I am left with 18 regions – the treatment region of Hedmark, and the remaining 17 regions constituting the donor pool from which a counterfactual will be created. In Section 5.7, Oppland is analyzed as a later exposure case in order to assess the program's generalizability.

Descriptive statistics for the estimation sample are given in Table A.2 in the Appendix in Supplementary material.

5. Results

5.1. Main results

Fig. 1 shows the sickness absence rate time series for Hedmark and its synthetic counterpart. The synthetic control region closely traces the actual one for the complete pre-treatment period, i.e. until and including the final quarter of 2011, even including the various seasonal cycles. The average difference in this period is 0.04 percentage points. The fit from the out-of-sample prediction provided in the validation period, i.e. all quarters of 2012 as well as the

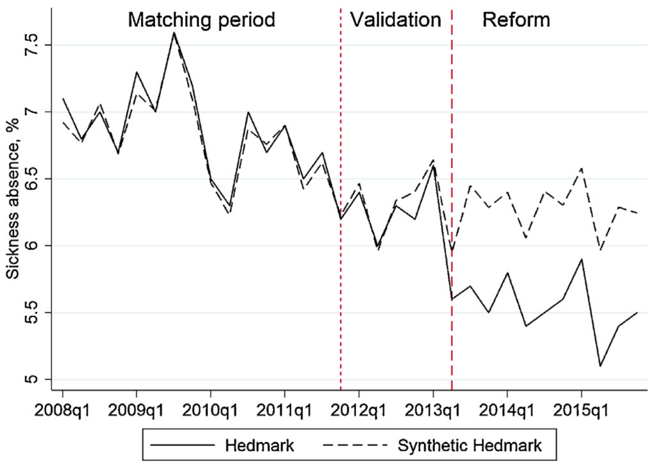

Fig. 1. Trends in the sickness absence rate in Hedmark and the synthetic control region.
Note: The dotted line at the fourth quarter of 2011 indicates the final quarter of the matching period. The dashed line at the second quarter of 2013 indicates the period in which the activation program was introduced.

Table 1
Means of predictor variables from the matching period.

	Treatment region		Donors		
	Hedmark		Nordland	Østfold	Telemark
Weight	Real	Synthetic	0.568	0.353	0.079
Sickness absence rate, %	6.844	6.800	6.694	7.063	6.388
Share of workforce					
with university education	0.256	0.257	0.257	0.256	0.259
in health and social services	0.219	0.216	0.224	0.203	0.218
in public sector	0.073	0.068	0.078	0.056	0.051
females	0.526	0.528	0.529	0.528	0.524
Share of inhabitants					
aged 20–29	0.104	0.112	0.113	0.110	0.113
aged 30–39	0.120	0.125	0.119	0.133	0.126
aged 40–49	0.142	0.144	0.143	0.147	0.142
aged 50–59	0.140	0.134	0.136	0.132	0.137
aged 60–69	0.122	0.113	0.113	0.113	0.115

first quarter of 2013, is also very tight, with an average difference of −0.06 percentage points, before the two graphs sharply diverge at the introduction of the reform.

A difference of around 0.4 percentage points occurs exactly in the quarter in which the activation program was introduced, drops further to around −0.75 percentage points in the following period, and remains at roughly that level for the rest of the time period. The difference for the whole post-treatment period is −0.72 percentage points, or −12% of the counterfactual post-treatment level, see Table A.3 in the Appendix in Supplementary material for details. One common worry with graded sick leave is that it may bring down the absenteeism rate in the short run, but increase it in the long run as people are not given the appropriate rest to recover. If this was the case, we should expect the estimated effect to decline and possibly reverse with time. It is reassuring that there is no sign that this is happening.[4]

The first columns of Table 1 show the values of the predictor variables for the treatment region – the real and synthetic Hedmark. To investigate whether there is a danger of interpolation bias, the final three columns list the values of the predictors for all the three regions receiving positive weights from the matching procedure (Nordland, Østfold, and Telemark). The weights are shown in the column headings. Nordland receives the largest weigh, 56.8%, and is particularly close to Hedmark on all variables. The other two regions are also quite similar, thus interpolating from very different regions is not a problem. A map of the regions is provided in Fig. A.1 in the Appendix in Supplementary material.

Table 2
Difference-in-differences estimates of the treatment effect on the absence rate. Percentage points.

	(1)	(2)	(3)	(4)
T	-0.606^{***}	-0.606^{***}	-1.645	-1.645^{***}
	(0.062)	(0.027)	(0.894)	(0.121)
T x graded share			2.024	2.024^{***}
			(1.683)	(0.215)
Cluster level	m	r	m	r
y mean, %	5.7	5.7	5.7	5.7
N	12,352	12,352	12,352	12,352

Note: Standard errors in parentheses. Cluster level indicates municipality (m) or region (r). All specifications include municipality fixed effects, time fixed effects and socioeconomic controls (the share of inhabitants with a university or college education, the share of the workforce employed in the health/social sector and the public sector, the female share of the workforce and the age composition of the inhabitants).
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

As a complement to the preceding analysis, results from difference-in-differences estimations at the municipality level are presented in Table 2 below. The point estimate from the baseline specification is around −0.6 percentage points, ref. Columns (1) and (2), thus somewhat lower than the synthetic control estimate. In light of the discussion of inference in settings with group-based treatments in Section 4, a central question here is how to handle the standard errors and interpret the uncertainty in these estimates. For clarity, the table presents standard errors clustered by both municipality and region. Column (1) shows the results when clustering by municipality, i.e. ignoring the serial correlation brought about by region-year shocks. This is likely to produce standard errors that are biased toward zero (Bertrand et al., 2004). Clustering by region in this extreme case of all treated units being concentrated in one group will likely make the problem of understated standard errors even worse (Conley and Taber, 2011). From Column (2), we can see that these standard errors are indeed even smaller.

Columns (3) and (4) show results from a specification with varying treatment intensity, expressed as the interaction of the treatment dummy with the pre-reform share of graded sickness absence certificates in the municipality. It is interesting to note that the positive sign on the interaction term indicates that the reform effect was largest in municipalities with low degree of grading before the reform. Are the standard errors valid in this case? Treatment intensity varies at the municipality level, hence we do not have the standard case of identical treatments within the group that calls for clustering at a higher level, but rather 22 differentially treated municipalities that can be seen as separate clusters. However, the fact that there will be a large, common reform component between these municipalities brings back the concern about serial correlation and understated standard errors. Thus, we must exercise caution in the interpretation of these estimates.

To get a sense of the economic significance of the estimated effect, we can compute the number of full-time equivalent working days the reform saved. Contracted hours in Hedmark in the post-treatment period was on average 4 000 000 days of work per quarter, thus a reduction in absenteeism of 0.72 percentage points constitutes close to 30 000 working days saved per quarter, or 120 000 days per year. Considering the fact that the replacement rate in the Norwegian system is 100 percent (up to a ceiling), it is clear that there are large savings involved for the public purse. How large?

[4] The results are largely the same when breaking the data down by gender, age and industry, see section A.1 in the appendix.

120 000 working days per year constitute around 520 working years. The average sickness benefit basis in 2014 was USD 47 375. Thus for one year, gross savings, in Hedmark only, are given by 520* USD 47 375 = USD 24.6 million. As there are around 79 500 employees in Hedmark, the savings constitute USD 310 per employee per year. Since sick pay is taxed at the same rate as regular income, this number is also the net savings. To the extent that these results might indicate lower permanent disability benefit uptake later on, the savings could magnify considerably.

The costs are more difficult to estimate. Even though a partly available employee in many cases will be preferable to no employee at all, in general, adapting to a worker with a graded sick leave certificate does carry costs for an employer. There are adaption costs also for the employee, however, long-term benefits associated with keeping a relation to the employer and the labor market in general, quicker recovery and avoiding earnings losses (Markussen, 2012) may go a long way towards counterbalancing or even exceeding these costs. Finally, the social security administration did incur some extra personnel costs related to the reform, but mostly handled the process by reallocating existing resources (NAV Hedmark, 2014). Thus on net, it is reasonable to assume that the reform was highly cost-effective.

5.2. Inference – “in-space placebo”

Even though the results above suggest that the reform had an effect on the absence rate, we do not know much about the uncertainty of the estimate in the absence of standard errors. To get something to compare the estimate to, we can reassign the treatment to the regions that did not undertake a reform and estimate a set of “placebo” reform effects. If this often produces estimated effects of a similar magnitude as above, we will lose confidence that the estimated effect for Hedmark is due to the reform. To get a sense of whether the estimated effect is particular to Hedmark, I subject all the 17 other regions from the donor pool to the same analysis, i.e. I construct synthetic counterfactuals in exactly the same way as above. Subsequently, I compute the difference between each actual region and its synthetic counterpart and plot all these difference-time series together to see whether Hedmark stands out. This plot is provided in Fig. 2, panel a), from which we can see that Hedmark is clearly extreme in that for no other region do we see such a large divergence between the actual and its synthetic control taking place exactly in the treatment period. The estimated effect for Hedmark is the largest one, thus the permutation based p-value based on all possible assignments of treatment is $1/18 \approx 0.056$.

Another way of evaluating the preceding results for all the regions is to compare the mean squared prediction error (MSPE) from the synthetic regions for the post- versus the pre-treatment period. A relatively small MSPE indicates a good fit, while a relatively large MSPE indicates a poor fit. Comparing the MSPE before and after thus takes into how well each particular region is matched in the first place when evaluating the post-treatment behavior. A large post/pre MSPE ration thus means that the synthetic region is not well matched after relative to before treatment, or in other words that there is a break in the fit. From Fig. 2, panel b), we see that Hedmark is clearly an outlier on this measure as well, with a post/pre MSPE ration of over 70, while the others are in the range between 0 and 12.

5.3. “In-time placebo”

The systematic placebo tests on all untreated units conducted above provides the main basis for inference with the synthetic control method. This can be thought of as falsification exercises in the spatial dimension. However, it is informative to undertake falsification tests also in the temporal dimension, since we clearly do not want our method to estimate a substantial intervention effect for a time when the intervention did not occur. To be able to do this requires data for quite a long period without any intervention. Fortunately, this is available in this case, thus I follow Abadie et al. (2015) and perform a so-called “in-time-placebo” study by reassigning the reform to the middle of the pretreatment period, which in this case is the third quarter of 2010, and rerunning the model. Fig. 3 shows that there is no effect from the placebo reform at this point in time. For further details, see Table A.4 in the Appendix in Supplementary material.

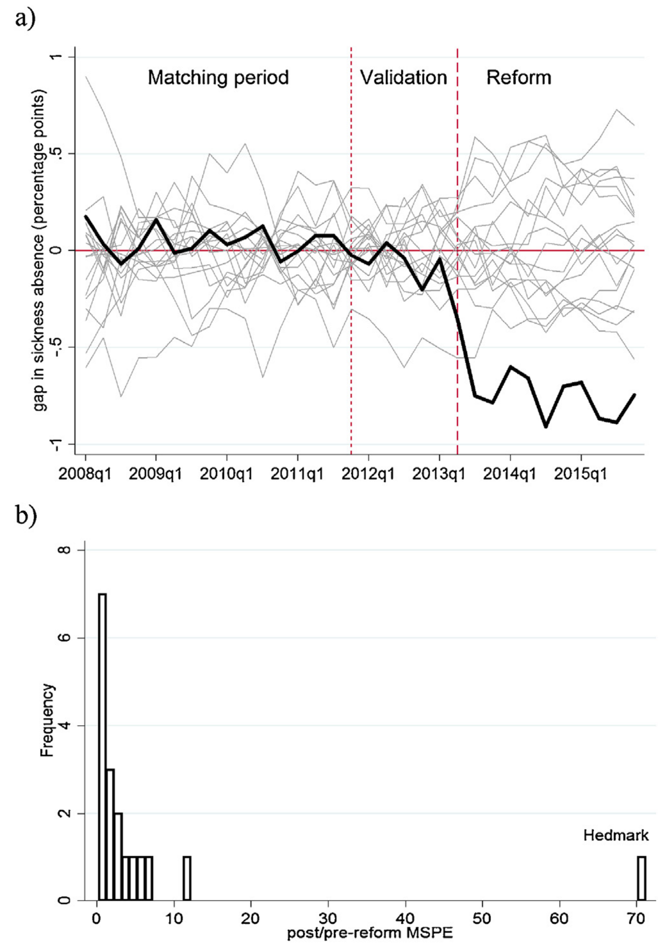

Fig. 2. a) Difference in sickness absence rate for all regions. b) Distribution of ratio of mean squared prediction error post vs. pre-treatment for all regions (MSPE after treatment / MSPE before treatment).
Note: The dotted line at the fourth quarter of 2011 indicates the final quarter of the matching period. The dashed line at the second quarter of 2013 indicates the period in which the activation program was introduced.

5.4. “Leave-one-out”

It would be worrying if the results were driven by particular donor regions. To check whether this is the case, this section presents results from successively leaving out each of the regions receiving positive weights in the baseline analysis and then performing the estimation from scratch. Fig. 4 shows that the results are robust to changes in the donor pool and do not rely on particular regions– the resulting counterfactual post-treatment paths are very close to the baseline estimate. For further details, see Table A.4 in the Appendix in Supplementary material. The results are also robust to how many and which covariates are used, as well as the length of the matching period.

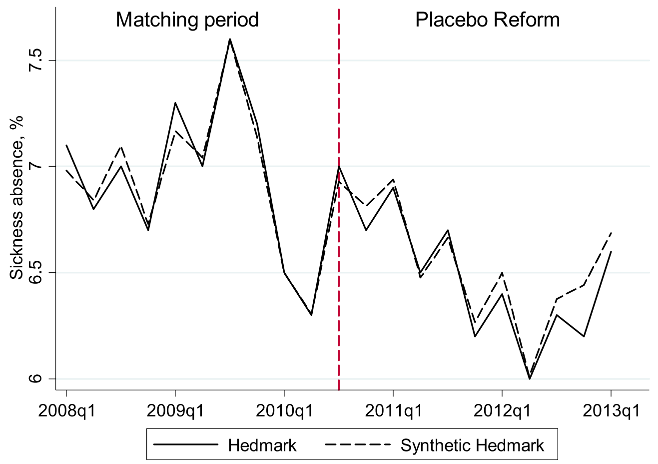

Fig. 3. In-time-placebo. Estimating effect of a placebo reform taking place in the middle of the pre-treatment period.
Note: The dashed line indicates the time of the placebo reform.

Table 3
Estimated effects on other outcomes. Percentage points.

	Short-term sickness absence rate (1)	Disability rate (2)	Employment rate (3)
Estimated effect	−0.0033	−0.200	−0.453
Pre- MSPE	0.0103	0.059	0.411
Post- MSPE	0.0044	0.113	0.486
MSPE ratio	0.435	1.919	1.18
p-value	0.944	0.667	0.611
Mean	1.07	10.3	68.5
N	576	486	576

Note: The socioeconomic characteristics employed are the same as in the baseline analysis (the share of inhabitants with a university or college education, the share of the workforce employed in the health or social sector and the public sector, the female share of the workforce and the age composition of the inhabitants). Data on short-term sick leave come from the quarterly labor force survey and are based on are self-certified absence spells, typically shorter than 4 days. Disability and employment rates are expressed as percent of the adult population. For disability, observations from after mid-2014 are excluded because Hedmark introduced a reform targeted at getting disability recipients back to work early that year (Kann and Lima, 2015). The estimated weights can be found in Table A.4 in the Appendix in Supplementary material.

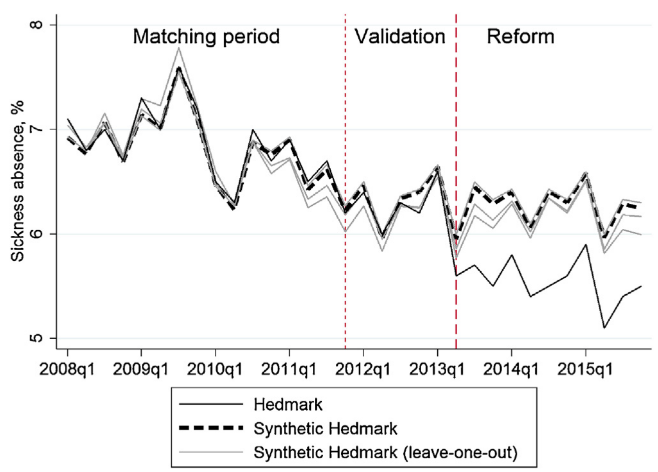

Fig. 4. Leave-one-out. Excluding each of the regions with positive weight in the baseline analysis.

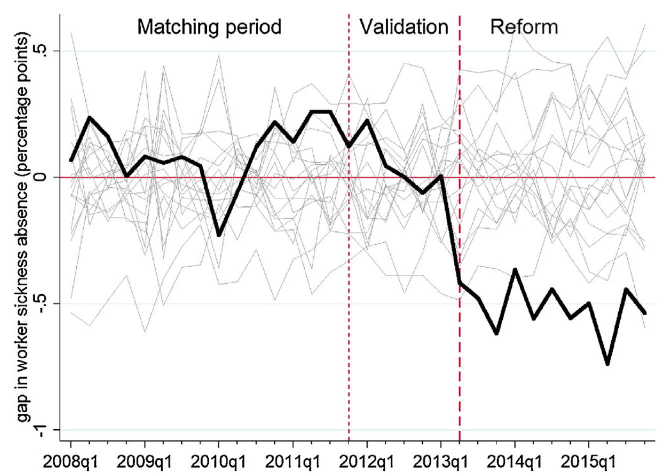

Fig. 5. Difference in the share of workers absent.
Note: The dotted line at the fourth quarter of 2011 indicates the final quarter of the matching period. The dashed line at the second quarter of 2013 indicates the period in which the activation program was introduced.

5.5. Other outcome margins

To get a more complete picture of the consequences of the reform, this section reports estimated effects on short-term sickness absence and disability and employment shares, see Table 3. The method and variables are exactly as in the main analysis in Section 5.1. The reform was targeted at long absence spells, as the activation requirement it set out to enforce only concerns spells of 8 weeks or more. It is therefore plausible to expect that the short-term absenteeism would be relatively unaffected. Consistent with this, the estimated effect on the short-term absence rate is very small, see Column (1) and has a high p-value. It is perhaps more likely that the reform could have had the unintended consequence of leading to people dropping out of the labor market, either to disability or to unemployment. Fortunately, there is little sign that this happened, though the uncertainty is considerable, see Columns (2) and (3). It should be mentioned that the program may also have prevented dropout from the labor market, by ensuring better follow-up of a sick-listed employee and making sure that the connection with the employer and the labor market in general is not lost through a long sickness period.

5.6. Mechanism

Unless an exception is given, the activation requirement will typically lead to a graded sick leave certificate, meaning that the employee will be partly back at work. There is an obvious mechanical relationship between the use of graded certificates and work hours lost due to absenteeism, since a person only partly absent will still perform the rest of their hours. This mechanical effect is important in its own right, but we would also like to know whether there are other effects going on, such as speedier return to full-time work.

First, instead of taking as the dependent variable the absence rate in terms overall work hours, we can analyze the share of workers absent, i.e. the number of workers with an ongoing absence spell (regardless of grading) divided by the total number of workers. This measure is a lot less stable, since it is impacted more strongly by seasonal variation in the number of short absence spells. It can nevertheless be used to answer the question of whether the reform simply increased the presence at work of people on leave, or whether there was also a reduction in the number of people with an absence spell (full or partial). The method and variables are exactly as in the main analysis in Section 5.1. Fig. 5 shows that as expected

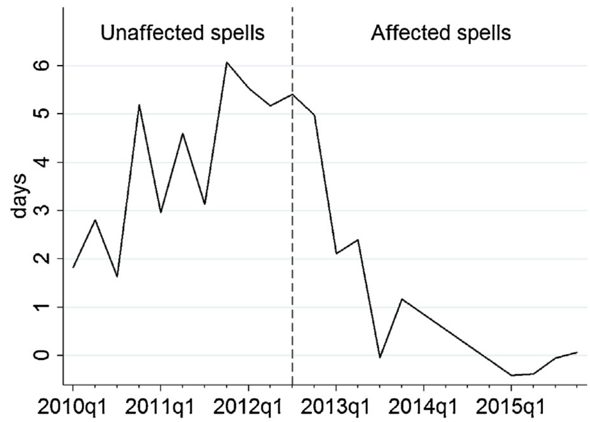

Fig. 6. Difference in average spell duration by spell start quarter.

the fit is not as good as with the absence rate, however it is clear that the reform also reduced the number of claimants. The reduction is approximately 0.5 percentage points, which implies that from the base of around 80 000 employees in the region, the reform lead to around 400 fewer persons with sick pay on a given day. The estimated weights can be found in Table A.4 in the Appendix in Supplementary material.

The fact that the formal activation requirement kicks in only after 8 weeks of absence makes it unlikely that the reform impacted the inflow to sick leave. To investigate whether accelerated termination of sick leave may have played a role, I will consider the duration of absence spells. The available duration data are average duration of completed spells by spell start quarter until the fourth quarter of 2013. However, since absence spells in Hedmark tended to be longer than in the rest of the country, it is hard to find a good match based on a convex combination of other regions. For that reason, the analysis will here diverge somewhat from what has been the case until now. From the baseline analysis of the absence rate above, I take the estimated weights presented in Table 2. Using these weights, I construct a synthetic Hedmark also for the average duration, then, knowing that the level is going to be somewhat off, subtract the values of the synthetic region from the real Hedmark to get the differences.

Fig. 6 shows the results. Since also spells that started before the reform, which took place in the second quarter of 2013, may have been influenced by the new regime, a dashed line is drawn three quarters earlier (at the third quarter of 2012) to distinguish spells not affected by the reform from spells that with increasing certainty would be affected. The maximum duration is 52 weeks, so very few spells that started more than three quarters before the reform would have been influenced by it. The graph shows that if anything, there was an increasing difference in average spell duration in the period before the reform, however there is a clear break at the point at which new spells begin to be affected. This is in line with what was found by Kann et al. (2014) in their examination the duration of new spells.

Thus the reform seems to have reduced the absence rate not only through exploiting the partial work capacity of temporary disabled workers, but also by speeding up the transition rate back to full-time work.

A final piece of evidence to consider is absenteeism broken down by diagnosis.[5] Musculoskeletal disorders and psychological disorders are by far the largest diagnosis groups, accounting for 40% and 20% of absenteeism, respectively. There is evidence that some degree of activity through work is both feasible and beneficial for

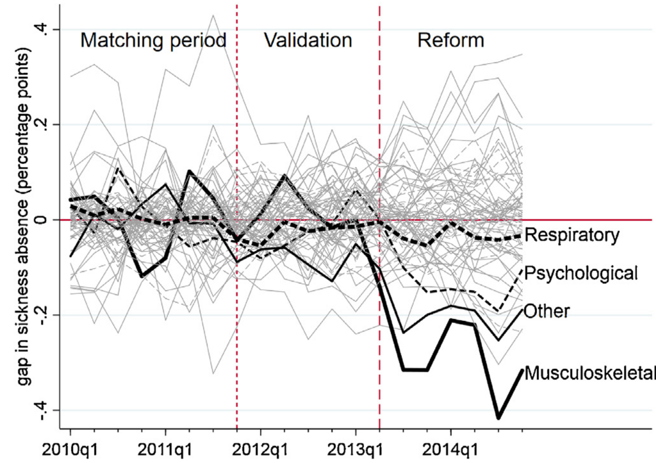

Fig. 7. Difference in absence rate by diagnosis.
Note: The dotted line at the fourth quarter of 2011 indicates the final quarter of the matching period. The dashed line at the second quarter of 2013 indicates the period in which the activation program was introduced.

many types of disorders (Waddell and Burton, 2006), and in particular for musculoskeletal disorders. In contrast, for respiratory disorders, which includes infectious diseases like the flu, presenteeism would be less desirable both from the point of view of recovery and out of concern for other workers. Thus, people suffering from respiratory diagnoses would typically not be the target of the activation requirement, and one would expect the reform to have much less of an effect on this type of absenteeism.

Fig. 7 presents estimates for these three diagnosis groups and a category of other diagnoses, constructed exactly as in the main analysis in Section 5.1. As expected, the largest decline occurred for absenteeism due to musculoskeletal disorders and the smallest, if any, for respiratory disorders, with diagnoses for psychological and other disorders in between. Weights and other statistics can be found in Table A.4 in the Appendix in Supplementary material.

5.7. *Later exposure case – the region of Oppland*

The region of Oppland has so far been excluded because it introduced a similar reform shortly after Hedmark and is thus badly suited as a control group. However, to analyze Oppland's reform as a separate case will be informative about the program's generalizability.[6] Some social security offices in Oppland started enforcing the activation requirement from the third quarter of 2013 – one quarter after Hedmark, and the rest followed later that year or early in 2014 (Kann et al., 2017). Subjecting Oppland to exactly the same analysis as Hedmark, this time with Hedmark excluded, shows that the reform was effective also in Oppland. The estimated effect is a decline in the sickness absence rate in the post-treatment period of 0.41 percentage points – substantial, though somewhat lower than the 0.72 percentage points estimated for Hedmark. As was the case with Hedmark, the estimated effect is the largest of all when subjecting the other regions to the same analysis, producing a permutation-based p-value of 0.056. The post/pre-MSPE ratio is 18.7, thus indicating a clear break in the fit. The results are shown graphically in Fig. A.4 and details are provided in Table A.4 in the Appendix in Supplementary material.

[5] Unfortunately, these data are only available up to and including 2014.

[6] I am grateful to the editor for this suggestion.

6. Conclusion

In this paper, I have shown that an activation program for workers on long-term sick leave in the Norwegian region of Hedmark reduced the absence rate by 12% compared to a synthetic control region created by a weighted average of similar regions. The results imply that exploiting the remaining work capacity of people on long-term sick leave has large potential for reducing unnecessary absenteeism and curbing costs related to social security transfers. Such an activation strategy represents an alternative to traditional attempts at welfare reform involving stricter screening or reductions in generosity, and may be more compatible with already existing legislation and contractual obligations, as well as easier to find support for across political priorities.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi: https://doi.org/10.1016/j.jhealeco.2018.09.007.

References

Abadie, A., Gardeazabal, J., 2003. The economic costs of conflict: e case study of the Basque country. Am. Econ. Rev. 93 (1), 113–132.

Abadie, A., Diamond, A., Hainmueller, J., 2010. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. J. Am. Stat. Assoc. 105 (490), 493–505.

Abadie, A., Diamond, A., Hainmueller, J., 2015. Comparative politics and the synthetic control method. Am. J. Polit. Sci. 59 (2), 495–510.

Andrén, D., 2014. Does part-time sick leave help individuals with mental disorders recover lost work capacity? J. Occup. Rehabil. 24, 344–360.

Andrén, D., Svensson, M., 2012. Part-time sick leave as a treatment method for individuals with musculoskeletal disorders. J. Occup. Rehabil. 22, 418–426.

Angrist, J., Pischke, J.S., 2009. Mostly harmless econometrics: an empiricists guide. Princeton University Press.

Barmby, Tim, Larguem, Makram, 2009. Coughs and sneezes spread diseases: an empirical study of absenteeism and infectious illness. J. Health Econ. 28 (5), 1012–1017, 2009.

Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? Quart. J. Econ. 119 (1), 249–275.

Brage, S.øren, Kristoffersen, Per, Lysø, Nina, 2014. «NAVs oppfølging av sykmeldte». Arbeid og Velferd, 1/2014, pp. 78–89.

Cameron, A.C., Miller, D.L., 2015. A practitioner's guide to cluster-robust inference. J. Human Res. 50 (2), 317–372.

Chatterji, Monojit, Tilley, Colin J., 2002. Sickness, absenteeism, presenteeism, and sick pay. Oxf. Econ. Pap. 54 (October (4)), 669–687.

Conley, T.G., Taber, C.R., 2011. Inference with "difference in differences" with a small number of policy changes. Rev. Econ. Stat. 1, 113–125.

Fevang, Elisabeth, Markussen, Simen, Røed, Knut, 2014. The sick pay trap. J. Labor Econ. 32 (2), 305–336.

Fiva, Jon H., Halse, Askill H., 2016. Local favoritism in at-large proportional representation systems. J. Public Econ. 143 (November), 15–26, 2016.

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 28, 2309–2316.

Hedmark, N.A.V., 2014. Sykefraværssatsingen i NAV Hedmark. Arbeid knyttet til aktivitetskravet for sykmeldte. Intern rapport, NAV Hedmark. [Internal report, NAV Hedmark.].

Høgelund, Jan, Holm, Anders, 2014. Worker adaptation and workplace accommodations after the onset of an illness. IZA J. Labor Policy 3, 17.

Høgelund, J., Holm, A., McIntosh, J., 2010. Does graded return-to-work improve sicklisted workers' chance of returning to regular working hours? J. Health Econ. 29, 158–169.

Høgelund, J., Holm, A., Eplov, L.F., 2012. The effect of part-time sick leave for employees with mental disorders. J. Ment. Health Policy Econ. 15 (4), 157–170.

Holm, Anders, Høgelund, Jan, Gørtz, Mette, Rasmussen, Kristin Storck, Houlberg, Helle Sofie B.øje, 2017. Employment effects of active labor market programs for sick-listed workers. J. Health Econ. 52, 33–44.

Johansson, P., Palme, M., 2002. Assessing the effect of public policy on worker absenteeism. J. Hum. Resour. 37 (2), 381–409.

Johansson, P., Palme, M., 2005. Moral hazard and sickness insurance. J. Public Econ. 89 (9–10), 1879–1890.

Kann, Inger Cahtrine, Lima, Ivar AndreasÅsland, 2015. Tiltak i Hedmark ga færre nye mottakere av arbeidsavklaringspenger, vol. 2. Arbeid og velferd, pp. 77–94.

Kann, Inger Cahtrine, Lima, Ivar AndreasÅsland, Kristoffersen, Per, 2014. Håndheving av aktivitetskravet i Hedmark har redusert sykefraværet, vol. 3. Arbeid og velferd, pp. 14–33.

Kann, Inger Cahtrine, Lima, Ivar AndreasÅsland, Brage, S.øren, 2017. Evaluering av forsøk med Hedmarksmodellen i Aust-Agder, Buskerud og Rogaland. NAV-rapport, vol. 3, 2017.

Kausto, J., Solovieva, S., Virta, L.J., Viikari-Juntura, E., 2012. Partial sick leave associated with disability pension: propensity score approach in a register-based cohort study. BMJ Open 2 (6), e001752, http://dx.doi.org/10.1136/bmjopen-2012-001752.

Kausto, J., Viikari-Juntura, E., Virta, L.J., Gould, R., Koskinen, A., Solovieva, S., 2014. Effectiveness of new legislation on partial sickness benefit on work participation: a quasi-experiment in Finland. BMJ Open 4 (12), e006685, http://dx.doi.org/10.1136/bmjopen-2014-006685.

Markussen, S., 2012. The individual cost of sick leave. J. Popul. Econ. 25 (4), 1287–1306.

Markussen, S., Røed, K., 2017. The market for paid sick leave. J. Health Econ. 55, 244–261, http://dx.doi.org/10.1016/j.jhealeco.2017.07.008.

Markussen, S., Røed, K., Røgeberg, O., Gaure, S., 2011. The anatomy of absenteeism. J. Health Econ. 30 (2), 277–292.

Markussen, S., Mykletun, A., Røed, K., 2012. The case for presenteeism—evidence from Norway's sickness insurance program. J. Public Econ. 96 (11), 959–972.

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. In: 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, 2010. Sickness, Disability and Work: Breaking the Barriers — Key Trends and Outcomes in Sickness and Disability. OECD, Paris.

Pichler, Stefan, Ziebarth, Nicolas R., 2017. The pros and cons of sick pay schemes: testing for contagious presenteeism and noncontagious absenteeism behavior. J. Public Econ. 156, 14–33.

Røed, K., 2012. Active social insurance. IZA J. Labor Policy 1 (1), 1.

Schneider, U., Linder, R., Verheyen, F., 2016. Long-term sick leave and the impact of a graded return-to-work program: evidence from Germany. Eur. J. Health Econ. 17, 629–643.

Shiri, R., Kausto, J., Martimo, R., Kaila-Kangas, L., Takala, E.-P., Viikari-Juntura, E., 2013. Health-related effects of early part-time sick leave due to musculoskeletal disorders: a randomized controlled trial. Scand. J. Work Environ. Health 39 (1), 37–45.

Van Vilsteren, M., Van Oostrom, S.H., de Vet, H.C.W., Franche, R.L., Boot, C.R., Anema, J.R., 2015. Workplace interventions to prevent work disability in workers on sick leave (review). Cochrane Database Syst. Rev. 2015 (10).

Viikari-Juntura, E., Kausto, J., Shiri, R., Kaila-Kangas, L., Takala, E.-P., Karppinen, J., Miranda, H., Luukkonen, R., Martimo, K.-P., 2012. Return to work after early parttime sick leave due to musculoskeletal disorders: a randomized controlled trial. Scand. J. Work Environ. Health 38 (2), 134–143.

Waddell, G., 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.

Ziebarth, Nicolas R., 2013. Long-term absenteeism and moral hazard—evidence from a natural experiment. Labour Econ. 24, 277–292.