

Gatekeeping versus monitoring: Evidence from a case with extended self-reporting of sickness absence.*

Arnstein Mykletun, Gaute Torsvik & Kjell Vaage

Abstract

This paper examines the impact of a policy reform that gave employees in one Norwegian municipality (Mandal) the opportunity to self-declare sickness absence for a whole year. To identify the effect of this reform we compare the development of absence in Mandal with the development in absence in similar municipalities. We apply both a standard difference-in-difference comparisons and the synthetic control method to quantify the effect (and uncertainty) of the reform. We find that allowing workers to bypass the physician as an absence certifier reduced sickness absence. The reduction in absence is especially strong for employees below 50 years. We find that it is the incidence of absence spells that declines, not their length. To explain the drop in absence we emphasize that the reform did not only remove the physician from the sick absence equation, it also put the employer more firmly into it, by prescribing a detailed follow up scheme (phone calls, meetings, flowers) for the employer (the first line-leader) and the employee calling in sick. A strict follow-up regime can be interpreted either as a concern for the workers well-being, or as monitoring. Interpreted in either way, the extended employer contact and activation of the worker calling inn sick may explain the drop in absence that we observe.

*We thank Edwin Leuven, Tarjei Havnes and Torben Mideksa for helpful comments.

1 Introduction

Many workers are insured against income losses due to temporary spells of illness. If health conditions drops below a threshold the worker is permitted to call in sick and obtain sick benefits as a compensation for the lost income. Sickness benefits can, as any type of insurance, be misused. Employees who are fit for work - with a health condition above the threshold and whose health will not decline by attending work - may choose to stay home and call in sick if the coverage rate is generous.

To constrain the moral hazard problem it is common to require a medical certificate from a physician in order to obtain sick benefits (OECD [2010]). Medical doctors are given the role as gatekeepers based on the notion that they are the best to judge whether an employee fulfills the requirements for sickness benefits or not. Their role as absence certifiers is, however, debated. If GPs consider themselves primarily as their patients advocate, requests for sick-leave certificates might be hard to overrule (Svårdsudd [2000], Carlsen and Nyborg [2009]). Their role as gatekeepers may also be weakened by their own economic interests as they may lose patients if they decline requests for a sick absence certificate. Another issue is that prescribing sickness certificates takes a toll on the time and energy of the physicians, resources that could - with extended self certification - be used on other patients. The conflicting motives that GPs have as the gatekeepers of health related welfare benefits and the fact that prescribing medical certificates for social benefits is costly, have spurred a several recent Scandinavian policy reforms and experiments that extend the period of self-reported sickness absence.

This paper examines the impact of a particularly radical experiment, a reform that gave workers the opportunity to bypass the physician as a sickness absence certifier. From Mai 2008, all municipality employees in Mandal municipality in Norway could self-declare sickness absence for a whole year. The rule elsewhere in Norway is that workers must obtain a medical certificate if they are absent for more than 3 or 8 days (depending on regime chosen by the firm). Extended self-certification will obviously reduce the resources medical doctors use to prescribe sick absence certificates. This is a desirable effect of the reform. The potential downside of extending workers discretion to self-declare whether or not they are

fit to attend work, is that it will increase level of absenteeism. Removing the physician as a gatekeeper may aggravate moral hazard problems in social insurance.

In a recent paper Markussen et al. [2013] exploit exogenous variation in the identity of the physician that patients meet. They find that the physician matter for the likelihood that a client obtains a medical certificate for sick leave. The fact that physicians matter for whether or not workers obtain sickness benefits indicate that they are, to some extent, gatekeepers to welfare benefits. In that case, we should expect absenteeism to increase after the Mandal reform. A randomized trial in Sweden fortifies this prediction (Hesselius et al. [2005]). Extended self-certification (from one to two weeks) was given to a randomized sample of workers in one urban and one rural region. A comparison with workers who did not obtain an extended self-reporting period show that the treatment group increased their absence significantly; on average absenteeism increased by 0.8 days per year (more for men).

Based on the results from the Swedish experiment we may expect a substantial increase in sickness absence in Mandal in the post reform period. There are, however, several reasons why a mechanical extrapolation of the Swedish results is unwarranted. The Mandal reform differs from the Swedish reform along many dimensions. First and foremost, in the generosity it grants employees to take responsibility for their decisions to report sick. The reform was branded “The Trust Project” and appealed openly to workers responsibility and reciprocity; “the reform grants you a lot of freedom and and we trust you will not misuse it”. Elsewhere, especially in the lab, it has been demonstrated that bestowing someone trust can enhance their trustworthiness (Falk and Kosfeld [2006]). To the extent that this mechanism kicks in, we might expect lower absence rates after the reform.

Furthermore, the reform did not only remove the physician from the sick absence equation, it also put the employer more firmly into it. The Trust Project prescribed a detailed follow up scheme (phone calls, meetings, counseling, sending cards and flowers) for the employer (the first line leader) and the employee calling in sick. The extended contact between employer and employee was conferred, in line with the “trust” branding of the reform, as a sign of the employer’s concern

for the workers well-being. It could, however, also be interpreted as enhanced monitoring, making it more costly to call in sick in boarder-line cases.

From a theoretical point of view the impact of allowing workers to bypass the physician as a sickness certifier is uncertain. Note, however, that even a zero effect on absence is interesting from a policy point of view, as the reform will reduce sick absence motivated visits to the doctor. Fewer consultations with the doctor saves public money (health consultations are subsidized by the state), and it also saves time and resources for the physicians. Hesselius et al. [2005] estimates that extending self declaration of absence with 7 days reduced the number of visits to the physicians with (a maximal estimate) 9 percentage points.¹ The potential gain in terms of reduced public reimbursement of physician costs is much larger in Mandal due to the size of the reform (going from 7 to 365 days of self-declaration).

To assess the impact of the reform on sickness absence, we compare the development in absence among municipality workers in Mandal before and after the reform with the change in absence among municipality employees in a number of other “similar” Norwegian municipalities. The development in absence “similar” municipalities is used to estimate of how absence in Mandal would have developed in the counterfactual case with no *Trust Project*. The critical assumption here is that the comparison municipalities have the same underlying time trend in absence among their employees as Mandal.

We use two methods to select comparison municipalities. First, we use a comprehensive list of municipalities that are classified in the same category as Mandal by Statistics Norway. This classification is based on the size of the municipalities and on their economic discretion. Second, we use the synthetic control method developed by Abadie and Gardeazabal [2003] and Abadie et al. [2010]. The essence of this method is to use the pre-reform period to construct a synthetic control unit (synthetic Mandal) - a convex combination of potential control municipalities - that resembles the treated unit along all dimensions that are important predictors for sickness absence in the post reform period.

Irrespective of the method we use to construct the comparison, the difference in

¹Potential, or maximal, because some individuals will visit the doctor even if they self-declare absence; the estimate made by Hesselius et al. [2005] is based on the assumption that none of those who self-report visits the doctor.

difference estimates show a considerable decline (15 - 20 %) in absence in Mandal in the post-reform period. A more detailed analysis of the data shows that the decline in absence is especially high (around 30%) for employees below 50. In a statistical sense our data consists of one treated unit (municipality workers in Mandal), which implies that we cannot use traditional inferential techniques to assess the uncertainty of the change in absence. In fact there is no sampling uncertainty in our case, since our data includes the whole population that was treated with the *Trust Project*. But it is still unclear - uncertain - to what extent the drop in absence observed in Mandal in the post reform period is due to the *Trust Project*, or if it can be explained by some other random event. One way to assess this uncertainty is to use some kind of permutation inference, that is, to pretend that the reform happened in other municipalities and then compare the impact of the placebo reforms on absence with the outcome in Mandal. When we do this, we find that the post reform development in Mandal stands out in the distribution of placebo-effects and hence we conclude that it is highly improbable that the drop in absence is due to a fluke of randomness.

In addition to the impact on absence, we find evidence that the reform induced older workers to stay longer in their jobs. Compared to its synthetic counterpart the fraction of employees above 50 years is considerable higher in Mandal. This effect may explain why the post reform drop in sickness absence among employees above 50 was modest. If the reform induced older workers with health problems to stay longer in their jobs, the absence rate among these workers may naturally increase.

The next session gives a brief introduction to the sickness insurance system in Norway and it also gives a description of the reform in Mandal. Thereafter we describe our data and the estimation method. Section 4 presents the results.

2 Institutional setting and the policy reform

2.1 Sickness absence in Norway

Sickness insurance is mandatory in Norway and covers all workers employed for more than four weeks. The compensation ratio is 100 percent from day one for a

maximum period of one year². Sickness benefits are paid by the employer for the first 16 days, and then by the National Insurance Administration (NIS) thereafter, for a maximum of 50 weeks. No medical certification is required for sickness spells lasting from one to three days, and self-reporting can be applied up to four times per year. As of 2001, firms are encouraged to join a publicly organized campaign, where one out of several components consists of allowing self-reported absence spells up to eight days, three times per year. Spells lasting more than three/eight days require a medical certificate, usually from a general practitioner. For spells lasting more than eight weeks an expanded certificate is required.

Due to the generosity of the coverage rate of sick benefits the level of absenteeism is high in Norway, around 7% of contracted work hours are lost due to sickness absence. Around 80% of the absenteeism is due to long term sickness spells (spells that last more than 16 days). The public expenditures associated with sickness absence is in the order of 2,5 % of GDP. Another problem is that those who obtain long-term sickness certificates entails a high risk of exiting the labour market and to receive social security benefits on a permanent basis Markussen et al. [2012].

2.2 Extended self-certification of sickness absence in Mandal.

There are 428 municipalities in Norway, the smallest has less than 300 inhabitants, the largest above 600000. Municipalities employ workers to provide goods and services. They are responsible for primary education (until 10th grade), outpatient health services, senior citizen services and municipal roads. Mandal municipality is above the average size of municipalities in Norway, in 2012 it had 15 000 inhabitants and 1200 employees. Due to the fact that all municipalities provide the same kind of goods and services the composition of their workforce, in terms of gender, age and education is quite similar.

Historically, the level of sickness absence for municipality employees in Mandal has been around the average for this sector in Norway. However, during the last decade the local authorities of Mandal stand out as being more than average

²There is an income ceiling of approximately 500,000 NOK p.a. (2012)).

active in their attempt to reduce sickness absence among their employees. In 2003, the municipality launched a “presence project” to reduce sickness absence among municipality employees. Arising from this project came an initiative towards the Ministry of Labour, requesting permission to “bypass” the physician as a sickness absence certifier. The suggestion was to let municipality employees, if they so wished, self-certify their sickness absence for the entire benefit period (one year). The conjecture was that employees would respond positively to the extended trust and counseling they would be given. In addition, the municipality administration proposed a detailed follow-up plan, where a strong involvement from the employees line managers would substitute for the GPs’ involvement and advice. For shorter spells, leaders were instructed to call the absentees (after 3 days and after 8 days). For longer spells the leaders were instructed to initiate a number of different meetings for individual counseling and follow-up plans, they should also regularly contact the absentee, send cards and flowers etc.³ All the initiatives popped up as outlook reminders on the mail system of the leaders, and if they did not follow up on the plan, a mail would be sent to the next leader in the hierarchy who then should take action. This system of a hierarchy of email based action reminders guaranteed that the follow-up plan was actually implemented.

The application of the system with extended self-certification of sickness absence was approved by the Ministry in June 2007. Mandal thereby became the first and only municipality – and firm – in Norway that was given permission to operate with a sickness insurance scheme that made the medical certificate from physician optional for the full length of the sickness spell. After some months of piloting a web-based system of self reported absence was in place in May 2008. In September 2008, almost 90 % of all sickness absence was self-reported. The project was prolonged in 2011, and it is still running.

3 Data

Statistics Norway divide municipalities into different classes based on their size and budget flexibility. Mandal is a medium sized municipality, a category which

³A detailed time plan for the initiatives taken toward the absent worker can be found in appendix A.

includes municipalities with between 5000 and 20000 inhabitants (Mandal had 15035 inhabitants in 2012). Municipalities are further divided into subgroups based on a set of economic variables. We have comprehensive information (a balanced panel) on sickness absence and other municipality statistics for 64 municipalities; Mandal and 63 municipalities that belong to the same category of medium sized municipalities. Although all of the municipalities in our sample are “medium sized” they vary in size, from having less than 500 to above 1500 employees.

Our data runs from 2003 until 2014. The first reform year is 2008, which means we have 5 observations before and 6 after the reform. For the reform year (2008) we only use the two last quarters, due to the fact that the reform was operative from first of May. We do not have individual data, but the information we have on contracted work days and lost days due to sickness absence is broken down by gender and by four age intervals, [16 – 39], [40 – 49], [50 – 59], and [60 – 69].

Our main analysis is based on absence spells that last longer than 16 days, mainly due to data availability and reliability. Our data is from the Social Security in Norway (NAV). Employers (municipalities in our case) are financially responsible for the short term absence, the state takes over the responsibility after 16 days. All employers have to report long-term absence to NAV in order to be reimbursed for sickness benefits that extend 16 days of absence. For short term spells only absence certified by a medical doctor is reported to NAV. Hence, for short term spells reported absence in Mandal will, if we use the NAV data, decline simply because only physician certified absence is recorded. Apart from data reliability, it is important to note that we capture the bulk of absence by measuring long-term absence; for municipality employees around 80 % of the contracted workdays that are lost due to sickness absence stems from sickness spells that extend 16 days. As a robustness check we also use short-term absence data we obtained from the employer organization representing the municipalities in our sample.

4 Results

4.1 Difference in Difference calculations

As a first take on the impact of the reform Table 1 compares before and after mean values for some key variables.

Table 1. Mean values before and after for Mandal and controls

	Mandal		Control ($N = 63$)	
	Before	After	Before	After
Employees	1032	1205	861	1006
Fraction of women	0.83	0.83	0.81	0.82
Fraction above 50	0.35	0.41	0.38	0.38
Sickness absence all	6.52	5.27	6.68	6.51
Sickness absence below 50	6.02	4.64	6.02	6.05

Table 1 reveals a relatively large drop in sickness absence in Mandal compared to the change in control municipalities. Absence is 1.25 percentage points lower in Mandal in the post-reform period while there is a slight drop of 0.17 percentage points in long term absence in the control municipalities. The differences in difference in means is 1.08 percentage points (17 % drop in long-term absence). The diff-in-diff is even larger for employees below 50 years, a 1.41 percentage points drop (22 %). There is also a difference in the development of the age-composition of the workforce; a 6 percentage point increase in the fraction of employees above 50 in Mandal, while there is no change in the fraction of workers above 50 in the control municipalities.

Comparing levels before and after the reform is indicative, but if we want to attribute the decline in absence to the reform we need to take a closer look of the development of sickness absence in Mandal and in the control municipalities. Figure 1 graphs the development in long-term sickness absence in Mandal and in the control municipalities. Each gray curve is a control municipality. The broken line is the mean of the controls. Panel a includes all employees and panel b show the development in absence for employees below 50.

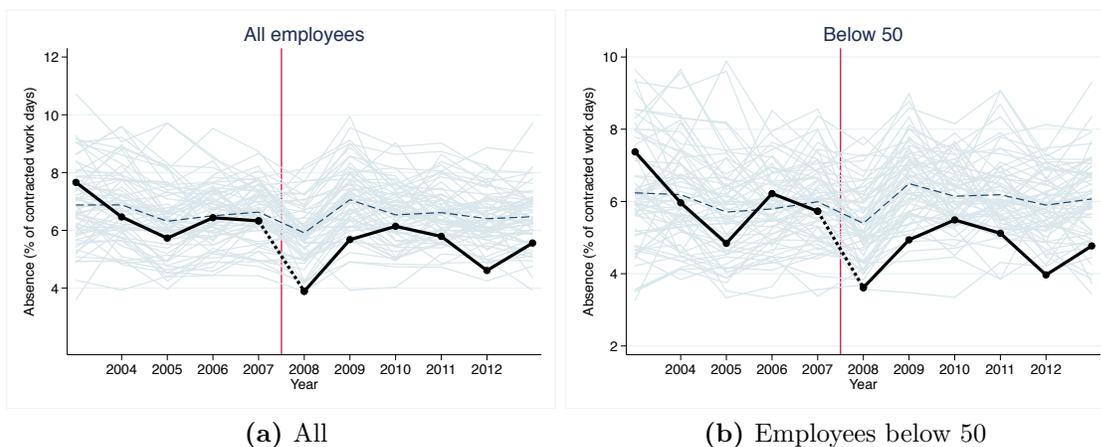


Figure 1. Evolution of long term absence before and after the reform

In Figure 1 absence in Mandal is at par with or above the mean of the absence levels in the controls before the reform and that it drops below the average in the post reform. The immediate drop (2008) is particularly large both if we take the average over all employees and for those below 50. For all employees the sharp drop in the first year of reform absence is followed by an increase in absence and absenteeism approaches the average in the control, before there is another low level of absence in 2012. For employees below 50 the absence level in Mandal stays well below the control for the whole post reform period. Another striking feature is the sharp decline in absence that came between 2003 and 2004.

Individual municipality curves show a substantial variation in absence over the years. The volatility is particularly high for relatively small municipalities (in small municipalities there is little data behind each observation). In Figure 2 panel a and b all municipalities with less than 800 employees are excluded (averaging over the data period Mandal has 1089 employees). With this restriction, we are then left with 33 control units.

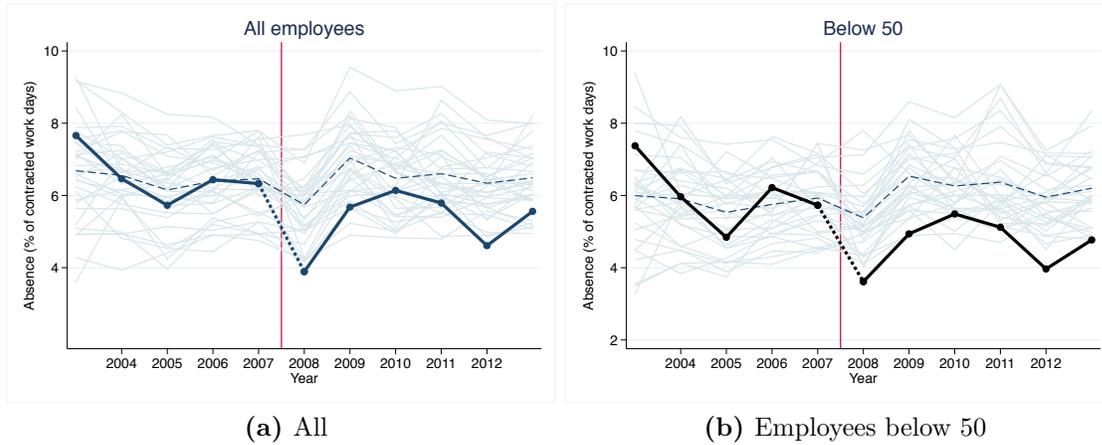


Figure 2. Evolution of long term absence before and after the reform with a restricted sample of controls.

The exclusion of municipalities with less than 800 employees does not change the mean of the control by much. The difference-in-difference between post- and pre-reform means are now -1.32 percentage points for all employees and -1.88 percentage points for employees below 50 (comparing with -1.08 and -1.41 if all municipalities are included). But there is considerably less volatility in the data.

To assess the statistical uncertainty of these estimates we use this restricted control sample. The simple idea here is to use a permutation test to find out whether the drop in absence observed in Mandal from 2008 and onwards stands out compared to the changes (after minus before) in other municipalities.

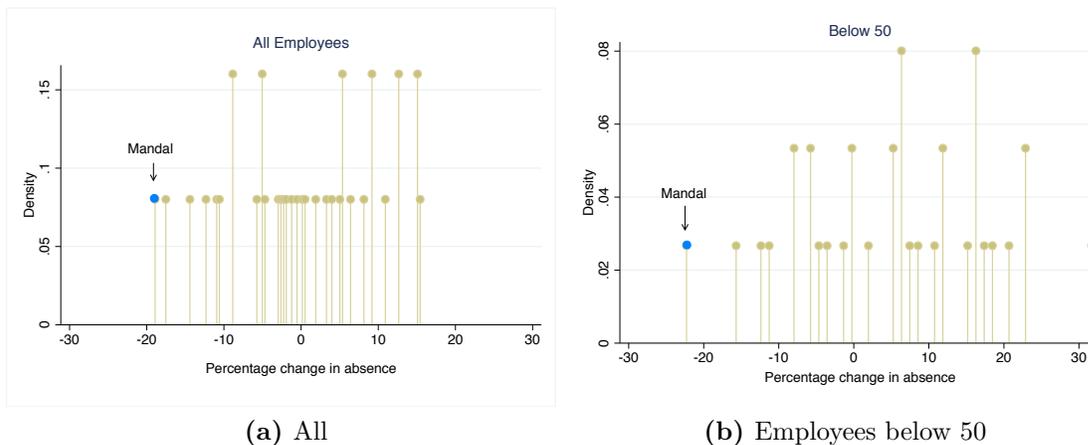


Figure 3. The distribution of relative changes in mean absence rates between pre- and the post-reform period

Figure 3 a and b show the distribution of changes in (post-reform average minus pre-reform average divided by the pre-reform absence) absence for all employees and for employees below 50. For all employees Mandal is tying with another municipality in having the most extreme percentage drop in absence. If we consider only workers that are below 50, Mandal is, by a large margin, the municipality with the largest drop in absence, but there is another municipality that has an even larger increase in absence.

4.2 Synthetic control method

Next we use the synthetic control method developed by Abadie and Gardeazabal [2003], Abadie et al. [2010] to estimate the causal impact of the reform. This method uses information from the pre-reform periods to construct a synthetic Mandal; a convex combination of the control municipalities that best resembles the treated unit - Mandal - prior to the reform. The idea is to match on all observable characteristics that may have an impact on the outcome of interest (sickness absence in our case). To account of unobservable variables that influence sickness absence it is recommended that one match on pre-reform values of the out-come variable (Abadie et al. [2010]). Once a synthetic control is constructed, the post reform difference in absence between the synthetic control and the reform

unit (synthetic Mandal and Mandal) gives a consistent estimate of the causal effect of the reform.

We match on lagged values of the dependent variable (sick absence), on the number employees, on the contracted work days divided by employees (in order to obtain a measure of part-time positions) and on unemployment rates in the different municipalities. Abadie et al (2010) suggest that one way to examine the uncertainty of the synthetic control estimates is to compare with placebo reforms in the control units (municipalities that did not grant their employees with 365 days of self certification). If the post-reform development in the treated unit is unusual compared to the placebo reforms this indicates that it is unlikely that the treatment effect is due to chance alone.

Figure 4 show the treated unit and the placebos, in panel a all employees are included, panel b compares employees below 50. The black thicker line is Mandal. We can see that the gap between Synthetic Mandal and Mandal is consistently negative, and it is in the lower range of the placebos (although there are a few “placebo gaps” that occasionally dips below the gap between the synthetic control and the treated unit).

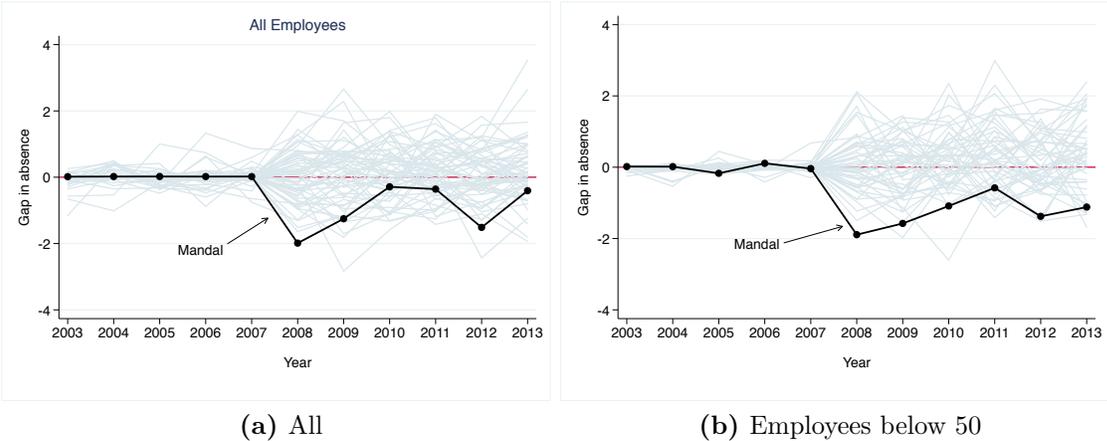


Figure 4. Gaps between synthetic X and X

Again it is interesting to calculate the average post-reform gap and depict the distribution of gaps and then check where in this distribution Mandal is located. When we do this we find that there are two (out of 63) that have a larger average

drop in absence in the post reform period if we compare with the synthetic counterpart. For below 50 Mandal is has, by some margin, the largest average drop in absence in the post reform period.

The general picture that emerges from the analysis above is that allowing workers to self-report sickness absence reduced long term absence, especially for employees below 50. The decline in absence endures over the post reform period (6 years), although the reform year stands out with a very large drop in absence.

5 Additional results

To better understand why extended self-certification and enhanced employer involvement lead to a drop in absence and why the drop was especially pronounced for workers below 50, we have analyzed the data in more detail. Markussen et al. [2012] argue that a nationwide Norwegian inclusive work-place reform in 2004 that, among other things, encouraged the substitution of graded for non-graded sick leave certificates, lead to shorter spells of sickness absence; a switch towards partial sickness insurance forced workers to utilize their remaining work capacity and this lead to a faster recovery of workers health and hence to a reduction sick benefits claims. Normally it is the physician, together with the employer and the worker, who determines the grading of absence spells. In Mandal, in the post-reform period the employer (the line leader) and worker decided the grading of the absence. Could it be that the Mandal reform increased the use of grading, which then reduced the length of the sickness spells, as found in Markussen et al. [2012]. Figure 5 show that this is not the case.

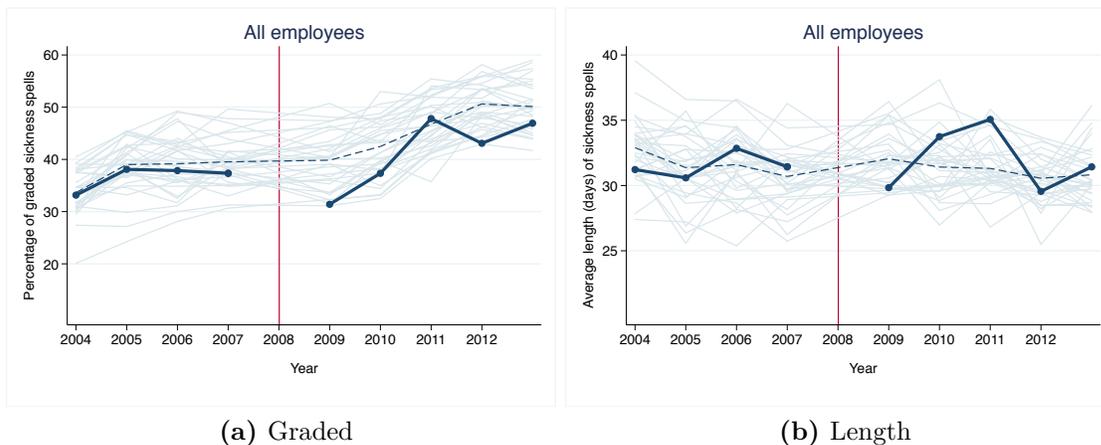
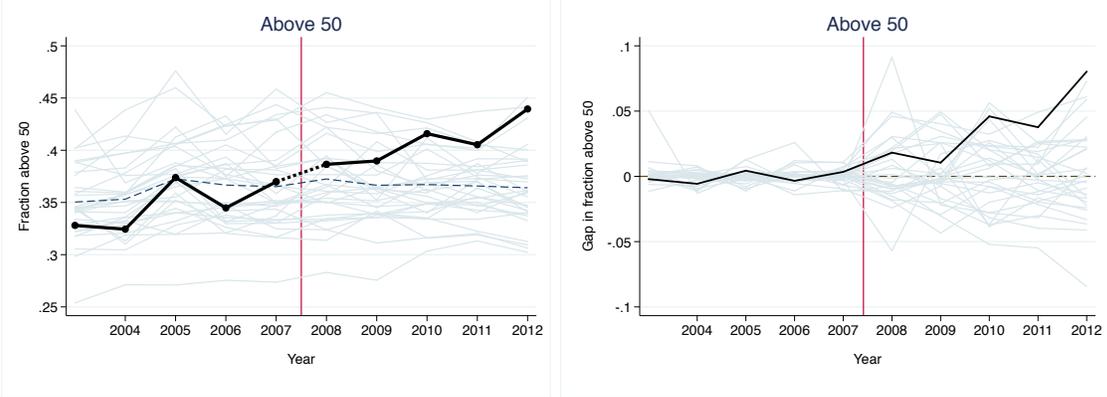


Figure 5. Grading and length of absence

It is interesting that the length of the spells basically follows the development in other municipalities. This means that the overall drop in absence that we observe in Mandal in the post reform period is due to fewer long-term spells of sickness absence. From this we can deduce that it is not all the activities that are initiated by the employer that improve the health and work capacity of the workers and bring them faster back to work. Rather, it is the prospect of being steered, activated and monitored by the line leader that enhances the barrier of entering a long-term spell of absence. But even if we can deduce this far, it is still hard, based on the data we have, to separate if it is the positive enhanced trust explanation or the more negative enhanced monitoring aspect of the reform that reduced the incidence of long-term absence. If it is the trust part of the reform, the extended discretion and responsibility individual workers are given to assess their own health related work capacity, that is particularly salient the reform will increase the work motivation of reciprocity inclined workers and reduce their inclination to be absent from work. If it is the monitoring and activation aspect of the reform that is salient, it will increase the psychological costs of reporting in sick and also reduce the incidence of absenteeism. It is difficult to separate these mechanisms.

Another interesting pattern in the data that can be discerned in Table 1 is that the fraction of workers above 50 increases in Mandal compared with the other municipalities. Figure 6 gives a more detailed account of the development

in this fraction. In panel a the fraction of workers above 50 is plotted against all other municipalities and their average, in panel b we have used a synthetic control method to let the data find municipalities with the same pre-reform (or pseudo reform for other municipalities than Mandal) trend.



(a) Fraction above 50

(b) Gaps between synthetic X and X

Figure 6. Fraction of workers above 50

Compared to the control municipalities there is an increase in the fraction of the municipality workers that are above 50 in Mandal in the period after the reform. This is perhaps not so surprising given that one of the most commonly travelled routes out of the labour force starts with a long-term sickness absence spell which then leads to disability insurance or some other form of permanent welfare support. The Mandal reform lowers the frequency of long-term absence spells which then later on induces a higher fraction of the workers to stay in their jobs.

6 Conclusion

In this paper we have examined the effect of an experiment carried out in a Norwegian municipality to reduce the sickness absence of the municipality’s employees. The core of the experiment is that sickness certification entitlement is transferred from physicians to the employees for the entire period of absence, which in Norway means a period of one year, maximum. It was also a part of the deal, however,

that the employer launched a strict follow up scheme. It is a matter of taste whether this should be interpreted as increased concern for the workers' well-being or increased monitoring. Our data cannot discriminate between these two interpretations of the employer's role.

We do, however, identify a significant drop in the sickness absence in Mandal after the introduction of the experiment, especially for employees below 50 years of age. This finding is robust across estimation methods (difference-in-difference and synthetic control). Another finding is that workers above the age of 50 seem to be induced to stay longer in their jobs, which in turn may explain the lack of reduction in sickness absence for this group.

An important conclusion from the experiment is that sickness certification can be taken off the hands of the physicians without a subsequent rise in sickness absence. Clearly, this has the potential of reducing the physician costs for the society. The Mandal experiment was not random and the exact mechanisms that lead to the reduced absence cannot be fully explored with our data and design. Hence, further research is necessary before we can recommend this scheme to be implemented on a national basis.

References

- Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. *American economic review*, pages 113–132, 2003.
- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490), 2010.
- Benedicte Carlsen and Karine Nyborg. The gate is open: Primary care physicians as social security gatekeepers. Technical report, Memorandum//Department of Economics, University of Oslo, 2009.
- Armin Falk and Michael Kosfeld. The hidden costs of control. *The American economic review*, pages 1611–1630, 2006.

Patrik Hesselius, Per Johansson, and Laura Larsson. Monitoring sickness insurance claimants: evidence from a social experiment. Technical report, Working Paper, IFAU-Institute for Labour Market Policy Evaluation, 2005.

Simen Markussen, Arnstein Mykletun, and Knut Røed. The case for presenteeism. evidence from norway's sickness insurance program. *Journal of Public Economics*, 96(11):959–972, 2012.

Simen Markussen, Knut Røed, and Ole Røgeberg. The changing of the guards can family doctors contain worker absenteeism? *Journal of Health Economics*, (32):1230–1239, 2013.

OECD. *Sickness, Disability and Work: Breaking the Barriers; A Synthesis of Findings Across Oecd Countries*. OECD, 2010.

Kurt Svärdsudd, Lars Englund. Sick-listing habits among general practitioners in a swedish county. *Scandinavian journal of primary health care*, 18(2):81–86, 2000.