Accepted Manuscript

Labour Market Effects of Activating Sick-Listed Workers

Kai Rehwald, Michael Rosholm, Bénédicte Rouland

Highlights

- We report the results of a large-scale randomized controlled trial conducted in job centers.

- We find an overall unfavourable effect of activating sick-listed workers on subsequent labour market outcomes.

- We also compare the relative effectiveness of alternative activation strategies.

- We exploit exogenous variation in activation regimes as an instrument for participation in a specific activity.

- Partial sick leave substantially improve prospects while paramedical care and traditional activation can be harmful.

Labour Market Effects of Activating Sick-Listed Workers *

Kai REHWALD   Michael ROSHOLM[†]   Bénédicte ROULAND[‡§]

Abstract

Using data from a large-scale randomized controlled trial conducted in Danish job centers, this paper investigates the effects of activating sick-listed workers on subsequent labour market outcomes. Comparing treated and controls, we find an overall unfavourable effect on subsequent labour market outcomes. Using variations in activation regimes (both between job centers and between randomly assigned treatment and control groups within a given job center) as an instrument for participating in a specific activity, we compare the relative effectiveness of alternative activation strategies. Our results show that the use of partial sick leave increases the length of time spent in regular employment, and also reduces the time spent in unemployment and in early retirement. Traditional active labour market programs and the use of paramedical care appear to have no effect at all, or even an adverse effect.

JEL Classification: J68, C93, I18

Keywords: Long-term Sickness; Activation Strategy; Treatment Effects; Field Experiment

1 Introduction

While most OECD countries provide sickness insurance (Heymann et al., 2010) and sickness poli-
cies are rapidly moving to center stage in the economic policy agenda (OECD, 2010), budgetary
considerations are at stake. The expenditure on paid sick leave in OECD countries amounted
on average to 0.8 percent of GDP in 2007.[1] Although this figure might seem rather low, it is
nevertheless a matter of great concern in the current context of growing public deficits and debt
burdens. In comparison, public spending on unemployment benefits reached "only" 0.55 percent
of GDP in the same year.[2] Furthermore, absence due to sickness also implies reduced labour
supply, lost production, and health-related costs.[3]

Beyond the financial aspect of paid sick leave, the reintegration of sick-listed workers into
the labour market is also a matter of great concern. Empirical research on the labour market
has shown that frequent and/or long-term spells of absence are associated with a higher risk
of unemployment (Hesselius, 2007), and can significantly reduce a worker's subsequent earnings
or prospect of employment (Markussen, 2012). The probability that a worker will then become
inactive and dependent on a permanent disability pension also increases.

The importance of well conceived sickness policies is clear in this context. Sickness policies
have recently shifted from being passive towards a more employment-orientated approach, thus
resembling the labour market policies for unemployed workers. The aim of such a shift is to
counteract the moral hazard problem in social insurance (see Røed (2012) for an enlightening
discussion on activation strategies). In particular, activating sick-listed workers should both
reduce benefit dependency and increase rates of employment.[4] Taking Denmark as an example,
activation measures were implemented in 16 percent of all periods of sickness benefit receipt in
the years 2009-2011, compared to only seven percent in the period 2005-2007 (Boll et al., 2010).
Activation measures for sick-listed workers include traditional active labour market programs
(e.g., internships), paramedical care (e.g., physical therapy), and graded return-to-work (i.e.,
partial sick leave).

Our aim in this paper is twofold. First, we wish to assess the effects of a stronger activation
strategy on sick-listed workers' subsequent labour market outcomes. Second, we aim to compare
the relative effectiveness of alternative activation strategies. Specifically, we use results from a

---

[1]OECD data on social expenditure, taken from the OECD (2010) report on *Sickness, Disability and Work.*
The term 'sickness' refers to public and mandatory private paid sick leave programs (occupational injury and
other sickness-related daily allowances).

[2]OECD data on Labour Market Programmes, extracted from OECD data bank (http://stats.oecd.org/). As
for Denmark—the country under consideration in this paper—expenditure on paid sick leave amounted to 1.4
percent, while public spending on unemployment benefit reached 0.96 percent of GDP in 2007.

[3]According to the Danish Ministry of Employment, absence due to sickness (short and long term) in 2006
reduced the supply of labour by five percent, which implies a cost of more than two percent of GDP.

[4]Refer to OECD (2010) for an outline of the main trends in recent reforms across the OECD.

large-scale randomized controlled experiment conducted in Danish job centers in 2009 among newly registered sick-listed workers. The treatment lasted four months and consisted of a combination of weekly meetings with caseworkers and intensive mandatory return-to-work activities in the form of graded return-to-work (partial sick leave), traditional activation, and/or paramedical care.

Our empirical strategy and key results can be summarized as follows. We first rely on a simple difference-in-means approach to identify the causal effect of participating in more activation requirements—whichever type—on subsequent labour market outcomes for newly sick-listed workers. Specifically, we estimate a causal *intention-to-treat* effect on several outcome variables, namely the accumulated weeks in regular employment, self-sufficiency (i.e., all forms of non-reliance on benefits), sickness, unemployment, early retirement, disability and *fleksjob*. Second, in the spirit of Markussen and Røed (2014), we exploit variations in activation regimes, both between job centers and between treatment and control groups within a given job center, as an instrument for participating in a specific activity to compare the relative effectiveness of the alternative activation measures.

Our findings reveal firstly that the experimental intervention as a whole has been ineffective. Sick-listed workers initially assigned to the treatment group spent less time in regular employment and self-sufficiency than their peers in the control group. Nevertheless, our results based on IV also show that a greater emphasis on offering graded return-to-work programs is associated with an increase in regular employment and self-sufficiency, and lower unemployment and early retirement. On the other hand, traditional activation and paramedical care appear to have either no impact at all, or even an adverse impact. Taken together, our results suggest that programs focusing on graded return-to-work are the most effective in improving sick-listed workers' subsequent labour outcomes. These programs are associated with strong and long-lasting effects, but only for workers sick-listed from regular employment and for those with physical (non-mental) disorders.

In line with the rich literature on the effectiveness of active labour market policies for unemployed workers (see Card et al. (2010) for a meta-analysis), our study relates to the expanding literature on the impacts of return-to-work policies for (long-term) sick-listed workers.[5] Return-to-work can be associated with various forms of interventions, including workplace-based[6], edu-

---

[5] There is also another branch of the literature that relates to the impacts of return-to-work policies for temporary disabled workers. See for instance Aakvik et al. (2005) and Markussen and Røed (2014) for a study of the Norwegian Vocational Rehabilitation program. While there is no absolute definition of long-term sick leave, workers typically call on a temporary disability insurance system following a period of sick pay (which is more generous than the disability insurance); however this is available only for a limited period of time (usually around one year).

[6] Reviewing recent medical research, Van Oostrom et al. (2009) concluded that workplace interventions are effective in reducing sickness absence among workers with musculoskeletal disorders compared with normal forms of care, although they are not effective in improving health outcomes.

cational, medical, and social interventions. The results are mixed, however; Frölich et al. (2004) for example, found that rehabilitation programs for the long-term sick (more than four weeks) in Sweden had no favorable effects at all, but that workplace interventions were less damaging than the alternative strategies. In a randomized study of the inflow of Swedish sick-listed individuals, Engström et al. (2015) found some negative effects associated with having early meetings to assess individuals' work capacity (more sickness absence and a higher probability of receiving disability benefits). In contrast, Everhardt and de Jong (2011) found strong positive impacts of return-to-work activities for long-term (nine months) sick employees in the Netherlands in terms of their likelihood of returning to work.

Some of the literature on workplace-based interventions focuses specifically on the effects of graded return-to work programs, i.e., some combination of part-time work and sickness benefits.[7] While accurate and reliable evidence remains scarce, Markussen et al. (2012) provides an exception. Using data collected from Norwegian administrative registers, the authors concluded that the use of graded (partial) rather than non-graded (full) sickness absence certificates reduces the length of periods of absence, and significantly improves the propensity for employment in subsequent years. Interestingly enough, Kools and Koning (2018) find that work resumption can be achieved faster when graded return-to-work is started earlier or at a higher rate of initial work resumption. Andrén and Svensson (2012) found that Swedish employees with musculoskeletal disorders assigned to part-time sick leave were more likely to recover to full work capacity than those assigned to full-time sick leave. From a randomized controlled trial performed in Finland among 63 patients with musculoskeletal disorders, Viikari-Juntura et al. (2012) showed that part-time sick leave reduced both the time taken to return to regular duties and the amount of sickness absence in the one-year follow-up period. In the Danish graded return-to-work program, Høgelund et al. (2010) found that participation in such a program significantly increased the probability that sick-listed workers returned to regular working hours. However, Nielsen et al. (2014) showed that its effect on the return to self-support differed substantially among the municipalities, and therefore warned against generalizing the results of the study to other Danish municipalities. Moreover, Høgelund et al. (2012) found no impact of the Danish graded return-to-work program for workers with mental health problems.

Based on a large-scale experimental design, the present study adds to the existing literature by offering a comprehensive evaluation of intensive mandatory return-to-work activities (activation requirements). In particular, we focus not only on workplace-based interventions but also on paramedical care, and thus compare the relative effectiveness of alternative intensive interventions

---

[7]Partial sick leave and partial sickness benefits are currently available in Sweden, Norway, Denmark, and Finland. The authorities have strongly promoted the use of these in recommending partial sick leave as the mechanism of choice, where sick leave is needed. See Kausto et al. (2008) for a review of the use of partial sick leaves in the Nordic countries. A similar arrangement has also been in place in the UK since 2010 (known here as the "Fit Note").

(traditional activation vs. paramedical care vs. partial sick leave). We also consider all kind of diseases rather than focusing on a specific subsample of sick-listed workers as others have done.

The rest of the paper is organized as follows. Section 2 includes details of the randomized experiment. Section 3 provides a description of the data and variables. In Section 4, we evaluate the treatment package as a whole, while we isolate the effects of each measure in Section 5. Some conclusions are offered in Section 6.

2 The randomized experiment

The Danish sick-leave policy. In Denmark, all employees, all self-employed, and all individuals receiving unemployment insurance benefits, are entitled to receive compensation for each day they cannot work due to sickness (whether the sickness is work-related or not), provided they have worked at least 120 hours within the thirteen successive weeks prior to their sickness absence.[8] Although the benefit period can be extended to more than a year under certain specific conditions, sickness benefits are normally available for a maximum of 52 weeks within an eighteen-month period. The employer finances the first 21 days of the sickness absence[9], while municipalities are responsible for funding the remaining period. The replacement ratio is up to 90% and benefits are capped at the maximum unemployment insurance benefits.

The municipalities play a key role in the return-to-work process. Besides managing sickness absence and work rehabilitation, it is their responsibility to monitor and assess recipients of sickness benefits. More precisely, municipalities must conduct an assessment of all sickness benefit cases no later than the eighth week of the sickness absence, and every fourth week from then on (or every eighth week in less complicated cases). Follow-up assessments must rely on updated and coordinated medical, social, and vocational information. The aim of these mandatory follow-up interviews is first to verify that the sick-listed individual is actually eligible for the benefit (i.e., he/she has a work incapacity) and second, to help him or her return to work as quickly as possible. Municipal case managers can implement various vocational rehabilitation measures, from job counselling to wage-subsidized job training and professional courses, and including graded return-to-work (partial sick leave).

It is worth noting that municipalities have economic incentives to reintroduce sick-listed individuals to the workplace because the state reimburses their expenditure on sickness benefit to varying degrees, depending on whether any return-to-work activities are implemented or not. Municipalities also have an incentive both to reduce expenditures on sickness benefit because

---

[8]Those conditions did not apply to unemployed workers whose eligibility for sickness benefits is conditional on being eligible for unemployment insurance benefits.

[9]At the time of the experiment (2009), the employer was required to finance the first fourteen days of the illness period; the rules on this changed in October 2009.

the entitlement to reimbursement only applies to cases lasting less than 52 weeks, and to use part-time (rather than full-time) sick-listing, because this also reduces the final burden on them.

Finally, if despite medical treatment and vocational rehabilitation the sick-listed worker is unable to return to ordinary employment, the municipality may refer him or her to a permanent wage-subsidized job (*fleksjob*) with reduced working hours and special tasks. To be eligible for a *fleksjob*, the sick-listed worker must have a permanently reduced work capacity of at least 50 percent and be no older than 65. The main difference between a *fleksjob* and graded return-to-work is that subsidized employment in a *fleksjob* is granted for an unlimited time while participation in a graded return-to-work program is always temporary. If the sick-listed individual cannot return to a *fleksjob*, the municipality may award disability benefits.

The experiment. In early 2009, the National Labor Market Authority in Denmark launched a randomized controlled trial (hereafter RCT) to test at a small scale some of the elements that were to be included in the forthcoming legislation on activation requirements in sickness insurance.[10] The overall purpose of this experiment was to examine whether sick-listed workers who participate in more activation requirements during their sickness period can achieve a greater degree of autonomy (in terms of returning to work and staying in work) than they would have done had they not behaved proactively. The present study is a report on the outcomes of this experiment.

The experiment was designed as an RCT conducted in 16 job centers across Denmark, with "random" assignment to treatment by birth year, even or odd. A total of 5,652 newly sick-listed workers were covered by the experiment, of which 2,795 were assigned to the control group and 2,857 to the treatment group. Figure 1 gives details on the timing of the intervention. In particular, individuals were notified that they had been assigned to the treatment group during the first follow-up meeting (no later than in the eighth week of the sickness absence) by means of a standard letter from The National Labor Market Authority.[11] Newly registered sick-listed workers born in odd years were subject to an intensive activation strategy (treatment group), while those born in even years were subject to the standard activation requirements (control group). Although assignment to treatment was predetermined, not random, we nevertheless consider our study to be an RCT; individuals were not made aware of the assignment mechanism, and a person's birth year is "random" from his/her own perspective.

Then, the treatment lasted for a maximum of four months and was meant to start between

---

[10]A new and stronger activation strategy was eventually rolled out nationally (with adjustments) in late 2009, when a new bill on the return-to-work of sick-listed workers was passed in Denmark.

[11]Participation in the experiment was therefore in reality conditional on being still sick-listed at that time. In particular, as shown in Table 2, the average elapsed duration of the sickness spell at the time of randomization was close to 8 weeks (7.8 weeks for the control group and 7.6 weeks for the treatment group).

1 week and 5 weeks after the first follow-up meeting with the caseworker.[12] It consisted of a combination of weekly meetings with caseworkers AND intensive mandatory return-to-work activities in the form of either a graded return-to-work (partial sick leave) and/or traditional activation and/or paramedical care. More precisely, *traditional activation* includes vocational guidance advice and courses aimed at enhancing skills, together with internships and on-the-job training. *Paramedical care* consists of courses on handling one's own situation, psychological consultations, nutritional counselling, and exercise including back exercises and other physical training. Lastly, the aim of the *graded return-to-work* measure is to support employees with reduced work ability to continue and return to work via partial sick leave. This involves working part-time (a reduction in working hours by at least four hours per week) and receiving a partial sickness benefit for the hours off work, on top of a partial salary (at a normal hourly rate). The underlying idea is that most sick leave days are a result of non-communicable diseases, and a person's work capacity while sick may be reduced but it is not zero. The return to regular working hours should take place as soon as possible, and certainly within the 52-weeks eligibility period of full sickness benefits. Graded return-to-work must be implemented with the agreement of the employer, the sick-listed worker, and the municipality. In practice, either the sick-listed worker and the employer arrange a graded return-to-work on their own initiative and then ask the municipality to approve this, or the municipality determines that the sick-listed worker is able to work part-time and therefore asks him/her to agree on a graded return-to-work with the employer. If the sick-listed worker refuses to enrol in the program even though the municipal case manager recommends it, the sick benefit can be reduced.

Within five weeks of the first interview, individuals in the treatment group were required to participate in some kind of activation (graded return-to-work, traditional activation and/or paramedical care) for at least ten hours a week for up to four months.[13] Table 1 provides details of the extent of these activation requirements. Compared with the control group, it is clear that the treatment group participated in more activities. For treated individuals, the number of return-to-work activities (all three types) was higher, and the interventions generally began earlier and lasted longer. For instance, the average week in which the first activity started fluctuated between 3.5 and 4.6 weeks for the treatment group while it was between 5.4 and 6.5 weeks for the control group.

Finally, it is worth noting that the job centers were responsible for the organization of the experiment. They carried out interviews and decided on the composition and content of any activation requirement, accounting for the individuals' needs and adapting to local conditions.

---

[12]Participants who returned to work before the end of the treatment period left the experiment.

[13]In an earlier version of this paper, we had a section on participants in the experiment who were excused from any kind of return-to-work activity. However, we are not dealing with it anymore as it is an endogenous outcome.

Figure 1: Time lines of the experiment

(a) treatment group
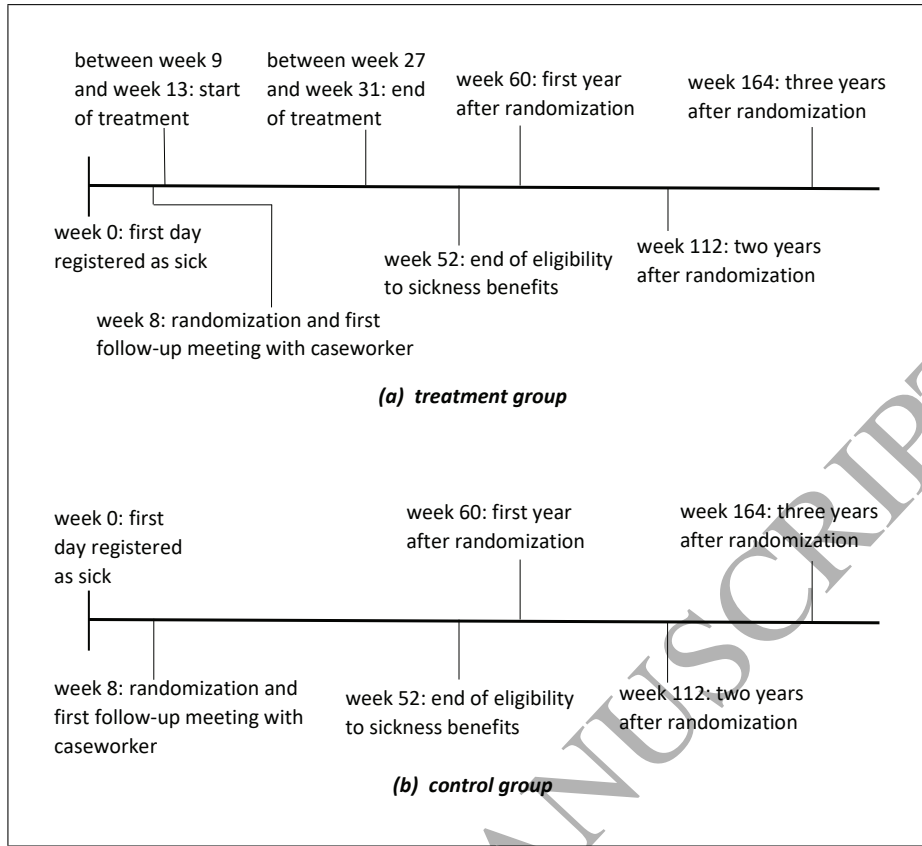
(b) control group

Table 1: Number of activation requirements by type and treatment status

Type	Control group (N=2,795)	Treatment group (N=2,857)
Panel A: *Traditional activation*		
Total number of weeks in traditional activation	1,111	6,129
Total number of traditional activation spells	97	512
... of which counselling and training	81	417
... of which on-the-job training	16	95
Avg. week in which the 1st traditional activation spell begins[a]	6.5	4.6
Avg. length of traditional activation spells (in weeks)[b]	11.9	12.5
Panel B: *Paramedical care*		
Total number of weeks in paramedical care	505	11,314
Total number of paramedical care spells	42	1,107
... of which courses in handling one's own situation	19	228
... of which psychologist	1	64
... of which nutritional counselling	0	63
... of which exercise	16	326
... of which back exercises or other physical training	4	166
... of which other	2	260
Avg. week in which the 1st paramedical care spell begins[a]	5.4	3.5
Avg. length of paramedical care spells (in weeks)[b]	12.6	14.4
Panel C: *Graded return-to-work*		
Number of initiated programs	366	566
Avg. week in which the program begins[a]	5.7	4.5

[a] Measured in weeks since first follow-up meeting with the municipal caseworker (typically eight weeks after the onset of illness).
[b] When calculating the average length of activation spells in weeks, uninterrupted sequences of alternative activities are treated as a single spell (93 spells in the control and 492 spells in the treatment group). Similarly for the average length of paramedical care spells in weeks (40 spells in the control and 787 spells in the treatment group).

9

From Figure 2, it is very clear that there is substantial variation in the use of the different activation measures, both between the 16 job centers covered by the experiment, and between treatment and control groups within a given job center.[14] As we explain in Section 5, we exploit this variation in "practice styles" to compare the effectiveness of the alternative activation requirements (graded return-to-work, traditional activation, and paramedical care).

Figure 2: Use of activation requirements by job center and treatment status
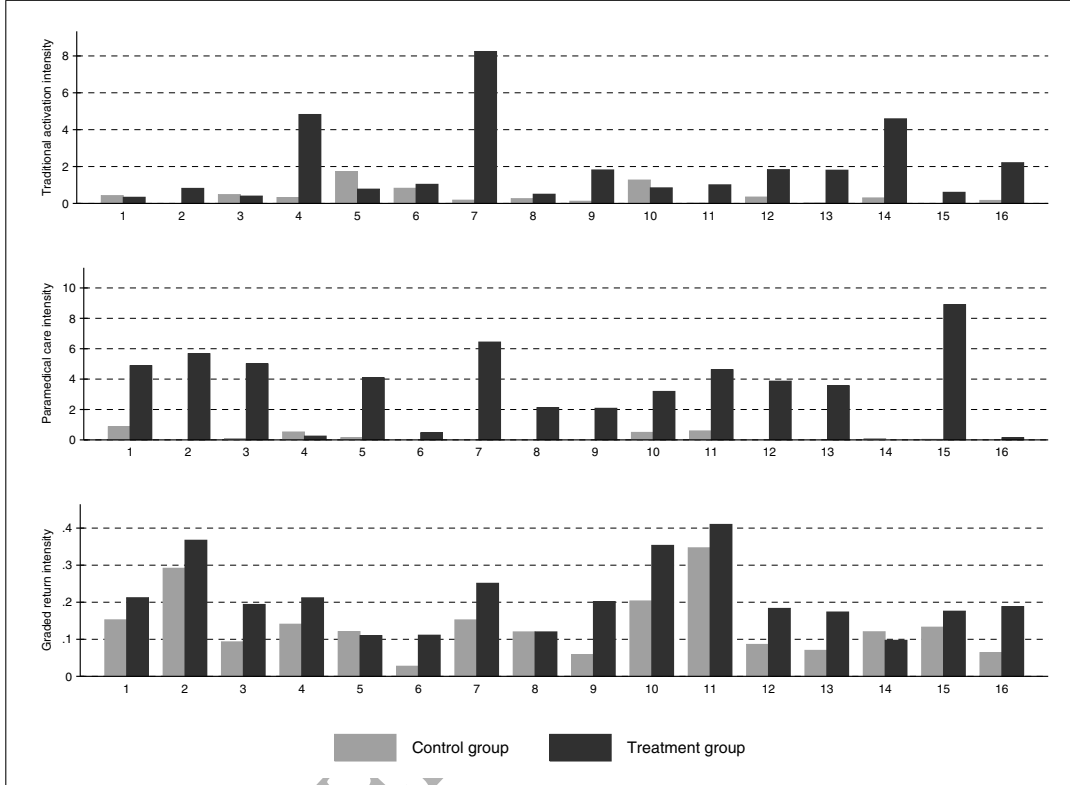
Notes: Indices 1 to 16 on the horizontal axes refer to job centers: (1) Bornholm, (2) Gentofte, (3) Greve, (4) København, (5) Ringsted, (6) Vordingborg, (7) Aalborg, (8) Morsø, (9) Randers, (10) Holstebro, (11) Herning, (12) Hors ens, (13) Svendborg, (14) Nyborg, (15) Odense and (16) Aabenraa. The intensity of traditional activation is calculated as the average number of traditional activation weeks per individual in the first 20 weeks after enrolment. Similarly for paramedical care. The intensity of the graded return-to-work programs is calculated as the proportion of individuals participating in a graded return-to-work program within the first 20 weeks.

3 Data and variables

The empirical analysis is based on the combination of four different data sets. First, we exploit unique Danish data derived from the controlled field experiment described above. These data include variables for each type of activation requirement (graded return, traditional activation and paramedical care) and for each meeting scheduled/held, as well as the number of hours per week in each type of activation and the timing of the intake. We can therefore follow participation accurately on a weekly basis. The data also include information about possible exemption from

---

[14]Similarly, the overall management of the standard activation requirements is regulated by law, but the assessment and the implementation of individual cases are controlled at job center level. Therefore, there is also substantial variation in the intensity of return-to-work activities among the control group.

activation.

Second, from the DREAM register we obtained weekly information about the type of social welfare benefit received[15], and about individual characteristics. DREAM is an amalgamation of information from several different sources, and is updated once per month by the The National Labor Market Authority.[16] Data from the DREAM register were obtained from the week in which the experiment began (i.e., between the first week of January 2009 and the third week of November 2009) up to the end of 2012, allowing a three-year follow-up. Specifically, we constructed seven outcome variables from the DREAM database, using the weekly information on social transfer payments, namely the cumulative number of weeks spent in regular employment (i.e., employed), in non-reliance on benefits (i.e., self-sufficiency), in sickness, in unemployment, in disability, in early retirement, and in *fleksjob*.[17]

Besides the treatment package, our explanatory variables include age, gender, marital status, country of origin (three categories—Denmark, western, and non-western countries), the state before sickness (three categories—job, unemployment benefits and self-sufficiency), the duration of elapsed sickness at the start of the experiment, the proportion of time spent on sickness payments (in each of the three years before the sickness), and the same proportion for time spent on public income support of any kind (also in each of the three years before sickness). Table 2, which shows means and standard deviations of the explanatory variables by treatment status, suggests that there are very few observed differences between the treatment and control groups. "Random" assignment based on birth year—even or odd—successfully balanced groups. The most significant difference relates to the status before sickness: the treatment group was on average in regular employment slightly less than the control group (76.3 vs. 79.9 percent), and in unemployment slightly more (17.4 against 14.2 percent).

Third, from Statistics Denmark we obtained data on socio-economic characteristics at the municipality level. In particular, for each municipality, we collected annual data on the total fertility rate, average age, and life expectancy for new born babies. We obtained quarterly information on the proportion of the working-age population with no more than ten years of

---

[15] The classification scheme includes all types of public income transfer schemes in Denmark. If a citizen does not receive social benefits in a given period, it is represented by empty week-variables.

[16] The week-variables are only allowed to contain one type of compensation code at a time. This implies that the types of social benefits are ranked. The ranking implies that if a citizen changes the type of social benefit in the middle of a week, only the highest ranked type is registered that week.

[17] More precisely, we first generated indicators for regular employment, sickness, unemployment, early retirement, disability and *fleksjobs* using the variable "status" from the DREAM register (i.e., the weekly information on labour market status). The outcome variable *regular employment* is defined using status 500 (i.e., the regular employment indicator is equal to one if status is equal to 500). Similarly, the sickness indicator is based on statuses 890 to 899, while the unemployment indicator is generated from statuses 111, 112, 124, 125, 130 to 138, 200 to 300, and 730 to 738. The early retirement dummy is made of statuses 621, 781 and 782. The disability dummy equals one if status is equal to 783. The *fleksjobs* indicator is based on statuses 622, 740-748 and 771-774. The *self-sufficiency* measure consists of statuses '-10', 500, 521, 651 and 652, which corresponds to the officially designated success criteria. Having defined these indicators, we then computed the cumulative number of weeks spent in each status.

Table 2: Pre-treatment characteristics by treatment status (N=5,652)

Variable	Control group		Treatment group		Difference
	Mean	Std. dev.	Mean	Std. dev.	
Panel A: *Socio-demographic characteristics*					
< 30 years old	0.161	(0.367)	0.144	(0.351)	-0.016*
30-39 years old	0.244	(0.429)	0.238	(0.426)	-0.006
40-49 years old	0.259	(0.438)	0.269	(0.444)	0.010
> 49 years old	0.337	(0.473)	0.349	(0.477)	0.012
Male	0.407	(0.491)	0.427	(0.495)	0.021
Married	0.597	(0.491)	0.581	(0.493)	-0.016
Danish origin	0.875	(0.330)	0.877	(0.328)	0.002
Western origin	0.046	(0.210)	0.047	(0.211)	0.001
Non-western origin	0.078	(0.269)	0.076	(0.264)	-0.003
Panel B: *Labor market history*					
Sick-listed from regular empl.	0.799	(0.401)	0.763	(0.426)	-0.036***
Sick-listed from UI-benefits	0.142	(0.350)	0.174	(0.379)	0.031***
Sick-listed from self-empl.	0.059	(0.236)	0.063	(0.244)	0.004
Elapsed sickness duration^a	7.841	(6.414)	7.621	(6.258)	-0.220
Time spent on sickness-ben.^{b}	0.188	(0.147)	0.188	(0.149)	-0.001
Time spent on sickness-ben.^{c}	0.075	(0.174)	0.076	(0.173)	0.001
Time spent on sickness-ben.^{d}	0.066	(0.164)	0.064	(0.162)	-0.003
Degree of pub. inc. support^{be}	0.302	(0.257)	0.311	(0.263)	0.009
Degree of pub. inc. support^{ce}	0.220	(0.320)	0.220	(0.321)	-0.001
Degree of pub. inc. support^{de}	0.244	(0.343)	0.243	(0.340)	-0.001
Panel C: *Diagnoses*					
Mental health disorder	0.300	(0.458)	0.292	(0.455)	-0.008
Non-mental disorder	0.7	(0.458)	0.708	(0.454)	0.008***
Number of observations	2,795		2,857		

^{a} at start of experiment (in weeks) ^{b} in year before sickness ^{c} in second last year before sickness ^{d} in third last year before sickness ^{e} any kind of public income support.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

schooling, on external and internal migration as a percentage of the population, and on the number of reported criminal offences per capita. As for the local labour market conditions, we calculated quarterly ratios as a percentage of the labour force for the working-age population outside the labor force, or receiving sickness benefits, or employed in the primary sector, as well as the number of full-time unemployed each quarter as a percentage of the labor force.

Finally, *jobindex.dk* is a collection of all the vacancies posted on the internet (online newspapers, job centers, job databases, etc.), which provides us with monthly information on the number of open vacancies and newly opened vacancies per unemployed person. We used this information to control for local environment characteristics when exploiting the variation in treatment strategies across job centers.

4   The effects of the treatment package as a whole

4.1   Empirical strategy

The first question we seek to answer is how participation in more activation requirements—whichever type—affects sick-listed workers' subsequent labour market outcomes, i.e., what are the effects of the treatment package as a whole? The field experiment with random assignment of participants to usual and intensified activation allows for a simple estimation strategy to study the causal effect of an intensification of return-to-work activities. Let us consider the following equation of interest:

$$Y_i = \beta_0 + \beta_Z Z_i + \beta_1 X_i + \varepsilon_i \tag{1}$$

where Y_i is the outcome of individual i, Z_i is a treatment status indicator equal to unity for individuals assigned to the treatment group (zero otherwise), and X_i is the vector of pre-treatment characteristics summarized in Table 2 (with age entering linearly). The coefficient β_Z corresponds to the (conditional) difference in the means of outcomes between treated and controls, and equals the *intention-to-treat* (ITT) effect $E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)$. Because treatment assignment is essentially random, $\widehat{\beta}_Z^{OLS}$ can be interpreted causally.

We consider seven measures for the dependent variable—number of weeks in regular employment, self-sufficiency, sickness, unemployment, disability, early retirement, and *fleksjob*—evaluated during the first, second, and third year after randomization.[18] The self-sufficiency measure is meant to cover all forms of non-benefit receipt. It encompasses individuals in regular (i.e., wage) employment, as well as self-employed, housewives and everyone else not receiving

---

[18]All individuals can be followed for 141 weeks, and outcome variables referring to the third year (up to 156 weeks) after enrolment are non-missing for all but eight individuals.

public income transfers.[19]

Additionally, in order to trace the trajectory of treatment effects over time in greater detail, let Y_i' be an alternative set of response variables denoting the cumulative total number of weeks—running sums counting from the week of intake—in each of the labour market statuses described above. We evaluate each of these outcomes at each week starting from the week of (individual) intake and ending with the 156th week after randomization (implying a total of 156 regressions per outcome variable). The results are presented graphically.

Besides its ease of use, the main advantage of a "difference in means" approach is that it produces informative estimates even in the presence of imperfect compliance on both sides. The main drawback, however, is that a simple comparison of average labour market outcomes between treated and non-treated only allows us to evaluate the treatment package *as a whole*.

4.2 Findings

We first present the effects of a more comprehensive activation package—consisting of intensified traditional activation, paramedical care, and graded return-to-work programs—on sick-listed workers' labour market prospects compared with those receiving treatment as usual. Table 3 shows conditional intention-to-treat (ITT) effects while Table A3 in the Appendix gives unconditional ITT effects (pure differences in means).[20]

It is immediately apparent from Table 3 that the experimental intervention as a whole has been ineffective. The combined treatment package had, on average, adverse impacts on subsequent labour market prospects. Regarding the outcome variables regular employment and self-sufficiency, estimated ITT effects are negative, moderate in size, and statistically significant during the first and second year after random assignment. The estimates suggest, for instance, that offering intensive rather than standard activation services reduced the time spent in self-sufficiency by one week on average during the first year, and one and a half weeks during the second year. The corresponding estimates for regular employment are slightly smaller in absolute terms. The impacts during the third year are also unfavorable, but end up being statistically insignificant once background characteristics are controlled for.

While the number of weeks spent in regular employment and other types of self-sufficiency has decreased for the treated, the number of weeks spent in sickness, disability and unemployment are not greater yet. The corresponding estimates are small in magnitude and not statistically significant at conventional levels across all years. Instead, the treated predominantly transited to early retirement and *fleksjobs*—subsidized jobs targeted at individuals with a permanently

---

[19]Exceptions are (subsidized) adult apprentices ("Voksenlærlinge") and individuals receiving state educational support ("Statens Uddannelsesstøtte"). Both groups are considered to be self-sufficient.

[20]As expected, we find that estimates in Tables 3 and Table A3 are rather similar, with standard errors being smaller in the former.

Table 3: Conditional intention-to-treat effects at different points in time after randomization

| Number of weeks in... | Time after randomization | | |
| | One year | Two years | Three years |
	ITT	ITT	ITT
Regular employment	-0.846*	-1.177**	-0.315
	(0.462)	(0.548)	(0.570)
Self-sufficiency	-1.081**	-1.549***	-0.851
	(0.475)	(0.557)	(0.575)
Sickness	0.039	0.104	-0.323
	(0.468)	(0.388)	(0.310)
Unemployment	0.281	0.398	0.296
	(0.275)	(0.382)	(0.388)
Early retirement	0.638***	0.815***	0.959***
	(0.111)	(0.220)	(0.252)
Disability	-0.079	0.009	0.281
	(0.109)	(0.226)	(0.308)
Fleksjob	0.245	0.635**	0.65*
	(0.228)	(0.285)	(0.346)
N	5,652	5,652	5,644

Notes: Pre-treatment variables include socio-demographic characteristics, individual labour market history and diagnoses. Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

reduced work capacity due to a medical condition. Participating in intensified rather than standard activation requirements increased the time spent in early retirement by about half a week during the first year after randomization, and by almost one week during the second and third year.

Figure 3 presents a magnified view of the trajectory of treatment effects over time and support previous findings. Each panel is based on 156 separate regressions, one for each week, of the cumulative total number of weeks in the respective labour market status on the treatment status dummy and the vector of background characteristics. The time series of treatment effects shown in Table 3 illustrates for example that, 156 weeks after random assignment, treated individuals spent on average two weeks less in regular employment than the non-treated. The adverse impact on non-benefit receipt is even more pronounced (around three weeks less). The figure also confirms that offering the treatment package promotes transitions into early retirement and *fleksjobs*. Considering these effects that accumulate in a sustained manner over time, one may conclude that the adverse impact of the treatment goes well beyond an initial locking-in effect. In contrast, cumulative total effects on sickness, disability and unemployment are small and not statistically different from zero over the entire domain.

In order to test for effect heterogeneity, we split the sample along several dimensions: we

15

Figure 3: Trajectory of treatment effects (cumulative total intention-to-treat effects)
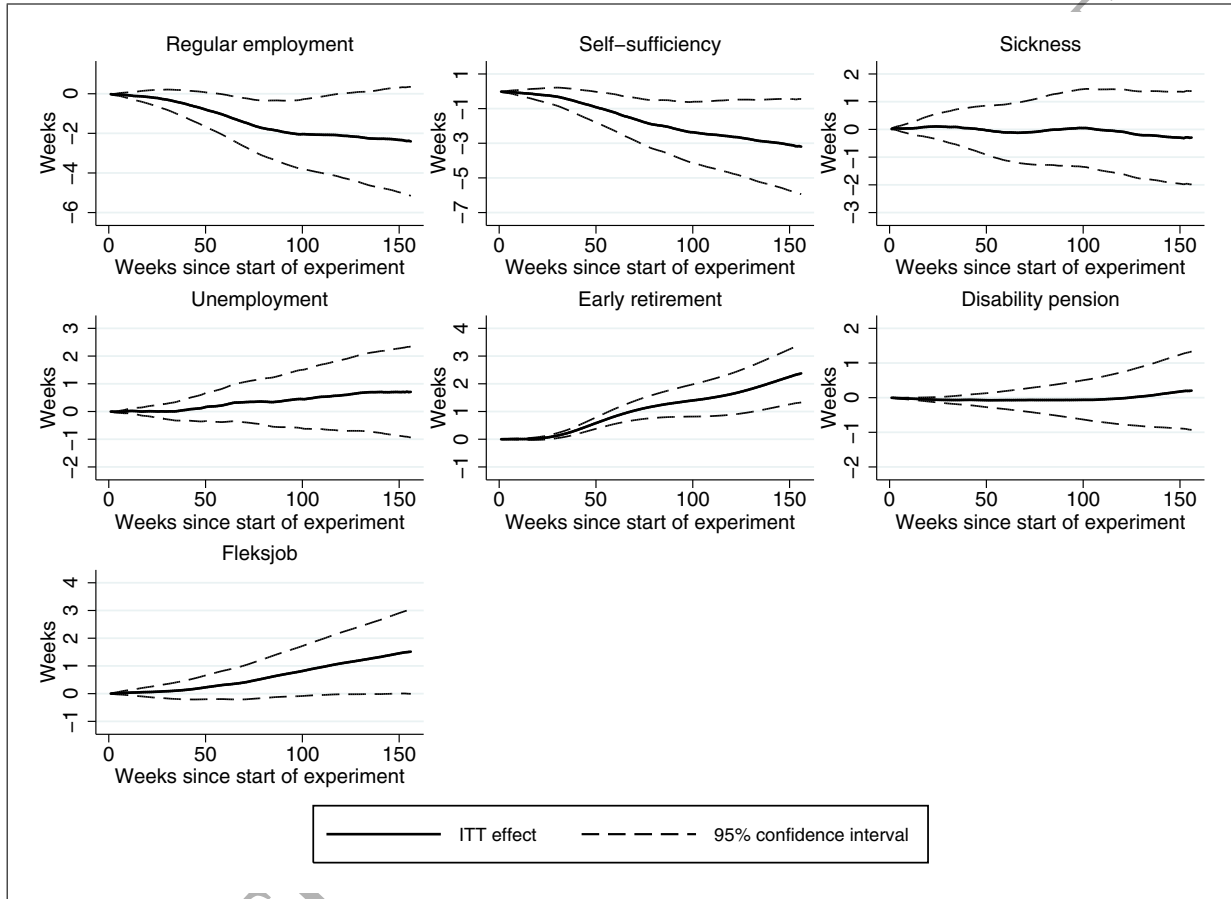
Notes: Each panel is based on 156 separate regressions (one for each week) of the cumulative total number of weeks in regular employment, self-sufficiency, sickness, unemployment, early retirement, disability and *flekjob* on a treatment status dummy and the vector of pre-treatment characteristics summarized in Table 2. The plotted intention-to-treat effects correspond to the difference in the average number of weeks spent in each status between treated and non-treated individuals (controlling for background characteristics) evaluated at a given point in time after randomization as indicated on the horizontal axes. The dashed lines depict the corresponding pointwise robust confidence intervals at the 95 percent level. The number of observations that each regression is based on varies between 5,652 for all weeks up to and including the 141st week after the start of the experiment (no missings) and 5,644 for the 156th week.

perform subanalyses by labour market status before sickness (regular employment vs. unemployment) and by diagnosis (mental vs. non-mental disorders). The results are shown in Tables A4 and A5 in the Appendix, and may be briefly summarized as follows. First, the treatment affects individuals sick-listed from regular employment and individuals sick-listed from unemployment in the same way (adverse impacts on regular employment self-sufficiency, *fleksjob*, no effect on sickness, disability and unemployment). Second, we did not find any different effects according to diagnosis.

In sum, participating in more activation requirements—whichever type—affects sick-listed workers' subsequent labour market outcomes unfavorably. The treated spent less time in regular employment and other types of self-sufficiency than their peers in the control group. The adverse impact is moderate in magnitude and goes well beyond initial locking-in effects. Lastly, the treatment promotes transitions into early retirement and *fleksjobs*, both of which are one-way tickets into indefinite periods of welfare dependency, and therefore dead-ends on the road to successful reintegration.

5 The effects of each activation measure included in the treatment package

5.1 Empirical strategy

The second question we seek to answer is how participation in each activation measure—traditional activation, paramedical care, graded return-to-work—affects sick-listed workers' subsequent labour market outcomes. Unbiased estimates can only be obtained if participation in an activity is as good as randomly assigned conditional on covariates.[21] Following the estimation strategy of Markussen and Røed (2014), we take advantage of the local variations in the implementation of the activation requirements, both between the 16 job centers involved in the experiment and between the treatment and control groups within a given job center (32 distinct "practice regimes"—see Figure 2), as a source of random variation in the probability of participating in a specific activity. Differences in the practice regimes may be seen in, e.g., the choice and combination of activation measures, the speed with which a newly registered client participates in an activity, and the length of the initiated activation spells.

The baseline equation of interest is:

$$Y_i = \alpha_0 + \alpha_P P_i + \alpha_1 X_i + \alpha_2 X_j + \varepsilon_i \tag{2}$$

---

[21]As the field experiment with random assignment of participants to usual and intensified activation was not designed as a multi-arm trial, the "difference in means approach" does not allow us to compare the relative effectiveness of the three alternative activities.

where Y_i is the outcome of individual i (we consider the same measures as in Equation (1)), X_i is a vector of individual pre-treatment characteristics summarized in Table A1 in the Appendix (where all non-binary variables, including age, enter in a quadratic way to allow for flexibility), and X_j a vector of municipality-level controls (socio-demographic characteristics and local labour market conditions; see Table A2 in the Appendix for a complete list).[22] P_i is a vector whose elements P_{A_i} are indicators of actual participation in an activity A (i.e, traditional activation, paramedical care, graded return-to-work). In particular, let P_{A_i} be equal to unity (zero otherwise) if individual i participated in activity A at some time during the first 20 weeks after enrolment—recall that the treatment period is meant to last only 18 weeks.

The parameter of interest α_P gives the average effect on y_i of participating in an activation measure. However, due to a selection problem, estimation of Equation (2) using OLS will provide biased results. To overcome this problem, we employ an instrumental variable for P_i to identify α_P. Our instrument is the practice regime the sick-listed worker is exposed to in the job center.[23] Since the "practice regime" is unobserved, we construct an estimate denoted PR_{A_i}.

To construct the "practice regimes" PR_{A_i} the sick-listed worker is exposed to, we first estimate a linear probability model explaining participation in the activation measures. In particular, we consider exits from a single state (*"sick and not activated"*) to multiple destinations: participation either in traditional activation, in paramedical care, or in a graded return-to-work program. Although a sick-listed individual can participate in a combination of alternative interventions over the course of his/her rehabilitation, our focus lies in the choice of first activity.[24] Let us consider the following model that we estimate separately for each of the three alternative activities:

$$P_{A_{ijd}} = \lambda_0 + \lambda_A D + \lambda_1 X_i + \lambda_2 X_{jd} + u_{A_{ijd}} \tag{3}$$

where $P_{A_{ijd}}$ is a destination-specific censoring variable equal to unity (zero otherwise) if sick-listed individual i, registered in job center j, makes a transition into activity A after having been unactivated for d weeks.[25] D denotes a vector of duration dummies (one for each week).

---

[22] Note that the vector of municipality-level characteristics is time-invariant, with the variables referring to the week of intake. This is because labour market outcomes arguably depend on *initial* conditions.

[23] Other studies have used similar instruments based on e.g. the practice style of physicians (Duggan (2005)), disability examiners (Maestas et al. (2013); Dahl et al. (2014)) and judges (Kling (2006); Aizer and Doyle (2015)).

[24] Accordingly, the data pattern used for estimation is characterized by single spells, one for each individual; repeat spells are ignored.

[25] Given these event indicators, we organize the dataset in the following way. First, because individuals who are not sick-listed in the week of intake—which is the case for 924 out of 5,652 experimental units—are not at risk of participating in any activity, they are excluded from the analysis beforehand; 4,728 individuals remain. Next, starting with a panel in person-week format, for each sick-listed worker we remove observations after the first participation into one of the three alternative activities. Uncompleted spells are right-censored in the absence of an event within the first 20 weeks or if the sickness spell ends. In short, the resulting panel is unbalanced and contains, for each sick-listed client, one observation per week at risk of being activated for the first time.

We then use the error terms from Equation (3) to construct PR_{A_i}.[26] To avoid that the unobserved individual characteristics of person i are included in the instrumental variable used to predict her participation into an activation measure, we remove the contribution of his/her error term, and we construct what is often referred to as a "leave-out mean"[27]:

$$PR_{Ai} = \frac{1}{n_j - 1} \sum_{k \in N_j^{-i}} \widehat{u}_{A_{kj}} \qquad (4)$$

where N_j^{-i} denotes the set of individual i's peers subject to the same practice style in job center j, and $n_j - 1$ is the cardinality of this set.

We perform 2SLS with Equation (4) as the first stage and Equation (2) as the second stage, with the goal of consistently estimating the parameter α_P. Under instrument validity, and assuming that there are no defiers (*monotonicity*), the coefficients α_P then identify local average treatment effects (LATEs), i.e., the average causal effects of actually participating in an activation requirement for the compliant subpopulation (Imbens and Angrist, 1994). We can also estimate the reduced form effect by directly regressing the following model:

$$Y_i = \zeta_0 + \zeta_{PR}PR_i + \zeta_1 X_i + \zeta_2 X_j + \varepsilon_i \qquad (5)$$

PR_i is the vector of "practice regimes" with elements PR_{A_i} (A=traditional activation, paramedical care, graded return-to-work). The coefficients ζ_{PR} identify the impacts of marginal changes in local treatment strategies on subsequent labour market outcomes and can thus be interpreted as *intention-to-treat* effects.

Instrument validity. Arguably, PR_{A_i} are valid instruments: they are *as good as randomly assigned* (i.e., *independent* of potential outcomes), *relevant* (i.e., partially correlated with the endogenous treatment receipt indicators) and *exogenous* (i.e., uncorrelated with the unobservable error term of the structural model).

Provided that there is no unaccounted for purposeful sorting of sick-listed workers into "practice regimes", exposure to a particular practice regime is as good as randomly assigned and thus independent of potential outcomes. It should also be clear that the proposed instruments are, by

---

[26]Markussen and Røed (2014: 16) argue that the residuals in model (3) have an appealing interpretation. Particularly, the sum of individual residuals, $\widehat{u}_{A_{ij}} = \sum_{d=1}^{D_{A_i}} \widehat{u}_{A_{ijd}}$ where D_{A_i} corresponds to the number of weeks a sick-listed worker i was at risk of making the transition into activity A, can "be interpreted as the estimated covariate-adjusted transition propensity at the claimant level". $\widehat{u}_{A_{ij}}$ is then equal to the (weighted) number of "lacked" waiting weeks for participation into activity A compared to what one would expect given the observed pre-treatment characteristics of client i and the municipality-level socio-economic characteristics of job center j. For instance, $\widehat{u}_{A_{ij}} > 0$ indicates that the transition happened earlier than expected.

[27]More precisely, we also exclude individual i when fitting the linear probability model specified in (3). Accordingly, we estimate 4,728 linear probability models per activity AS, excluding one individual at a time and computing one datapoint (PR_{A_i} for the excluded individual) per iteration.

construction, strongly correlated with the potentially endogenous vector of actual participation in activation measures. In addition, the *instrument relevance condition* can (and will) be tested.

Regarding the *exclusion restriction*, we need to maintain that the only channel through which practice styles affect labour market outcomes is through their effect on program participation. In particular, we need to assume that the instruments are uncorrelated with unobserved determinants of the outcome. In what follows, we will argue that this assumption is plausible. For this purpose, it is instructional to think of the error term in equation (2) as being composed of two parts. The first part contains unobserved determinants of labour market success that are specific to the individual sick-listed worker, i.e., unobserved individual characteristics such as ability, motivation, or the loss in work capacity due to sickness (which is not fully captured by controlling for diagnoses). The second part comprises all remaining factors, i.e., unobservables that are non-specific to a particular client. This second component consists first and foremost of (potentially unobserved) local labour market conditions and other municipality-level influences.

Now, note that, first, the set of instruments relevant for individual i is *completely* exogenous to individual i in the sense that neither individual i's characteristics nor individual i's treatment history have *any* impact on the instruments. Individual characteristics (observed or unobserved) should therefore be orthogonal to the instrument. It remains to be argued that the same is true for the second part of the error term. We need to maintain that the practice styles are uncorrelated with unobserved local labour market conditions and other municipality-level unobservables determining the outcome. If we think of practice styles as being the result of the interplay between national statutory provisions and a "combination of individual judgment, guesswork, personal experience, and convenience" (Markussen and Røed, 2014: 6), i.e., if practice styles are first and foremost determined by factors unrelated to current local conditions, this requirement is arguably satisfied. Given this line of reasoning, it should also be the case that *observable* municipality-level variables have no significant effect on observed treatment portfolios. To test for selection on observables, we regress the indicators activity intensity plotted in Figure 2 on the vector of municipality-level characteristics, X_j, as described in Table A2 in the Appendix. For traditional activation and paramedical care, none of the estimated coefficients is statistically significant at conventional levels. Moreover, F-statistics suggest that they are also jointly insignificant. For graded return-to-work, only one out of 16 covariates ends up being statistically significant (at the ten percent level). We take this as evidence in favor of the assumption that there is no selection based on (observable) municipality-level variables; this supports the exclusion restriction.

5.2 Main results

Table 4 shows the estimated effects of participating in each activation measure—traditional activation, paramedical care, graded return-to-work. Estimates may be summarized as follows. First, both traditional activation and paramedical care seem to have adverse impacts on subsequent labour market outcomes. Participation in a traditional activation program promotes unemployment during the first year after enrolment and is clearly not helping sick-listed workers to reintegrate into the regular labour market. Similarly, traditional activation programs are ineffective in reducing the degree of welfare dependency. The effects of being exposed to paramedical care are even more detrimental. Our results indicate that participating in such a program reduces the time spent in regular employment and self-sufficiency by about one and a half (6 weeks) to two months (8 weeks) during both the first and second year after intake. These effects are highly statistically significant. Paramedical care also tend to significantly increase the use of *flekskobs*.

In contrast, requiring that sick-listed workers participate in graded return-to-work programs seems to be a very successful strategy. Partial sick leaves increases the time spent in regular employment and non-benefit receipt substantially. At the same time, we find favorable effects on sickness during the first year, on unemployment during the second year, and on the use of *fleksjobs*.

Tables A7 and A8 in the Appendix present a decomposition of the effects of participating in some activation requirements, by labour market status before sickness (employed vs. unemployed) and by diagnosis (mental vs. non-mental). The general impression is that the results for the full sample (Table 4) are entirely driven by the large subsample of workers sick-listed from regular employment. Treatment effects for the subsample of unemployed individuals cannot be distinguished from zero (only one estimate is statistically significant at conventional levels). The finding that graded return-to-work is ineffective for sick-listed unemployed workers comes as no surprise, given that unemployed workers typically have no work to return to.

In contrast, graded return-to-work programs perform best for sick-listed workers with a non-mental condition. While the effects of these programs are generally positive, although imprecisely measured in some specifications, paramedical care and traditional activation appear to have either no or even adverse impacts. These findings correspond very well with the results for the full sample displayed (Table 4).

Taken together the empirical evidence strongly advocates the use of graded return-to-work programs for workers sick-listed from regular employment as a result of a non-mental medical condition. The picture for individuals with a psychiatric disorder is less clear. In fact, since none of the estimated effects exceeds its standard error, the results for this group are entirely

Table 4: Effects of participating in alternative activation requirements (LATE) evaluated at different points in time after randomization

	I Regular employment	II Self-sufficiency	III Sickness	IV Unemployment	V Early retirement	VI Fleksjob
	Dependent variable: Number of weeks in ...					
Panel A: *During 1st year after random assignment (N=4,728)*						
Traditional activation	-3.333	-2.529	-1.556	2.796*	0.403	1.066*
	(2.876)	(2.764)	(2.985)	(1.629)	(0.453)	(0.627)
Paramedical care	-7.084**	-6.551**	5.316*	-0.794	0.141	1.674***
	(2.761)	(2.654)	(2.866)	(1.564)	(0.434)	(0.602)
Graded return	30.290**	23.777**	-24.260*	1.425	0.639	-6.355**
	(11.980)	(11.515)	(12.435)	(6.787)	(1.885)	(2.610)
Panel B: *During 2nd year after random assignment (N=4,728)*						
Traditional activation	-4.620	-2.359	-1.539	3.396	0.116	1.672
	(3.252)	(3.174)	(2.386)	(2.410)	(1.262)	(1.427)
Paramedical care	-7.510**	-8.063***	2.458	3.399	0.790	3.416**
	(3.122)	(3.048)	(2.291)	(2.313)	(1.212)	(1.370)
Graded return	27.285**	17.971	-9.886	-17.532*	3.474	-10.530*
	(13.546)	(13.224)	(9.941)	(10.038)	(5.258)	(5.946)
Panel C: *During 3rd year after random assignment (N=4,720)*						
Traditional activation	-5.960*	-3.088	-0.377	0.830	0.350	1.110
	(3.374)	(3.256)	(1.859)	(2.361)	(1.801)	(1.898)
Paramedical care	-3.466	-3.606	-0.931	1.942	1.101	3.314*
	(3.255)	(3.140)	(1.793)	(2.277)	(1.737)	(1.831)
Graded return	25.984*	14.294	-0.722	-11.517	4.372	-10.294
	(14.172)	(13.674)	(7.808)	(9.916)	(7.564)	(7.973)

Notes: The table shows the estimated effects of participating in alternative activation requirements (LATE). Each panel is based on six separate IV/2SLS regressions (second stage). See notes to Table A6 for a description of the outcome variables.

Included instruments: see Table A1 and Table A2 in the Appendix; week of intake dummies.

Excluded instruments: vector of "practice regimes" (PS_i). F-test of excluded instruments (first stage): traditional activation: 228.75; paramedical care: 480.75; graded return: 20.28.

Standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

inconclusive.

Finally, Table A6 in the Appendix displays the average ITT effects of marginal changes in practice regimes, i.e., the results from the reduced form.[28] It comes at no surprise that the LATE estimates (LATE) tend to be greater in magnitude than these ITT effects. They also suggest that prioritizing graded return-to-work programs has favorable effects. It decreases the incidence of sickness during the first year after random assignment, and increases time spent in regular employment and non-benefit receipt. We find that the favorable effect on regular employment persists during the second and third year after enrolment. During these later years, the impacts on self-sufficiency and sickness are also beneficial, but not statistically significant. While the impact on early retirement cannot be distinguished from zero, prioritizing graded return-to-work programs reduces the likelihood of getting trapped into permanent welfare dependency by taking up subsidized employment under a *fleksjob* arrangement. In stark contrast, giving priority to traditional activation and paramedical care impairs subsequent performance. Both of these treatment types exert a negative impact on the incidence of regular employment and non-benefit receipt.

5.3 Robustness

As part of the IV estimation strategy, we argued that the vectors of practice regime characteristics provide legitimate instruments for the potentially endogenous participation in activation requirements. A key assumption is the *independence* of the instruments with potential outcomes: for the individual sick-listed worker, being exposed to a particular practice regime is *as good as randomly assigned*, and so are the instruments. As a matter of fact, much of the variation in the practice regimes is *completely random* because of the experimental intervention—the variation between treatment and control group *within* a given job center is truly random.

Given this setup, we consider three robustness tests. First, we isolate the truly random part of the variation in practice regimes. So far, a practice style has been defined to consist of a particular job center-treatment status cell: the control group in Copenhagen constitutes for instance *one* practice style. This definition allows us to exploit the variations in the use of activation measures, both *between* the 16 job centers covered by the experiment and between treatment and control group *within* a given job center (32 distinct practice regimes). Let us now consider only the truly random part of the variation in practice styles, by letting a practice style correspond to an individual's treatment status (which leads to the definition of two practice

---

[28]Following Markussen and Røed (2014: 19), we have normalized the vectors of practice styles by scaling its elements PS_{A_i} by the inverse of the absolute difference in the average value of PS_{A_i} between the local treatment regimes applying the respective treatment activity A least and most. Consequently, a unit difference corresponds to the difference described above and parameter estimates in Table A6 can be interpreted as the expected change in the outcome variable "resulting from a movement from the treatment environment giving lowest priority to the strategy under consideration to the one giving it highest priority".

styles). The advantage of this approach is that it rules out purposeful sorting into treatment regimes and thus guarantees *independence*. Table 5 shows the results of this robustness test. It is immediately apparent that our main findings (Table 4) are robust to this alternative approach. The results are the same as before, both in terms of the direction of the effects and their statistical significance (in fact, the estimates presented here are even more precise than those in Table 4). In particular, we again find that graded return-to-work programs are associated with favorable effects, while traditional activation and paramedical care perform poorly.[29]

Table 5: Effects of participating in alternative activation requirements (LATE) evaluated at different points in time after randomization, accounting only for truly random variation between treatment and control groups

| | Dependent variable: Number of weeks in ... | | | | | |
	I Regular employment	II Self-sufficiency	III Sickness	IV Unemployment	V Early retirement	VI Fleksjob
Panel A: *During 1st year after random assignment (N=4,728)*						
Traditional activation	-2.720** (1.114)	-2.706** (1.133)	-0.644 (1.195)	1.413** (0.703)	0.065 (0.197)	0.790*** (0.221)
Paramedical care	-3.529*** (1.262)	-4.032*** (1.283)	1.229 (1.353)	0.592 (0.796)	0.366 (0.223)	0.418* (0.250)
Graded return	10.606*** (0.899)	11.005*** (0.915)	-8.124*** (0.965)	-2.890*** (0.567)	-0.124 (0.159)	0.083 (0.178)
Panel B: *During 2nd year after random assignment (N=4,728)*						
Traditional activation	-2.126 (1.338)	-2.857** (1.364)	-1.332 (1.025)	0.119 (0.988)	0.211 (0.537)	2.180*** (0.561)
Paramedical care	-5.118*** (1.516)	-6.193*** (1.544)	1.811 (1.161)	1.077 (1.119)	1.242** (0.608)	0.871 (0.635)
Graded return	9.216*** (1.080)	8.701*** (1.101)	-3.355*** (0.827)	-3.496*** (0.798)	-1.576*** (0.434)	0.210 (0.453)
Panel C: *During 3rd year after random assignment (N=4,720)*						
Traditional activation	-1.178 (1.391)	-2.349* (1.406)	-0.502 (0.811)	-1.215 (1.010)	0.112 (0.768)	2.564*** (0.786)
Paramedical care	-2.833* (1.572)	-3.633** (1.590)	-0.613 (0.917)	1.251 (1.142)	2.445*** (0.868)	0.880 (0.889)
Graded return	8.091*** (1.122)	7.025*** (1.134)	-0.517 (0.654)	-3.519*** (0.814)	-2.368*** (0.619)	-0.146 (0.634)

Notes: First robustness test. The table shows the estimated effects of participating in alternative activation requirements (LATE). Each panel is based on six separate IV/2SLS regressions (second stage). See notes to Table A6 for a description of the outcome variables.
Included instruments: see Table A1 and Table A2 in the Appendix; week of intake dummies.
Excluded instruments: vector of local treatment strategy characteristics (PS_i). F test of excluded instruments (first stage): traditional activation: 2,007.29; paramedical care: 839.18; graded return: 2,837.23.
Standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As a second test of robustness, let us pretend that the experiment did not take place. In the absence of the experimental intervention, all sick-listed workers would have participated in the standard activation requirements. Therefore, for this second robustness test, we only use data

---

[29]Note that the statistics shown in Table 5 are of smaller magnitude than those reported in Table 4. It is worth noting that this does not cast doubt on the reliability of the estimates. By changing the way a local treatment environment is defined, we end up with a different set of instruments, which may be associated with a different compliant subpopulation and thus different LATEs (Angrist and Pischke, 2008).

from the control group, and we now focus on the variation in the use of activation requirements across 16 practice regimes. The results are shown in Table 6.The effects reported here tend to be in the same direction as our main results. However, few of the estimates are statistically significant at conventional levels. For one thing, the precision loss might be caused by the sample size reduction (given that we are now using only control group data). For another, it may arise from the reduced extent of exploitable variation in practice regimes. Given these considerations, one might interpret the results shown in Table 6 in the following way: while the direction of the effects is (by and large) accurately measured, the variation in practice regimes in the control group is too small to derive statistically significant estimates, given the relatively small sample size. Precision could be increased either by increasing the sample size (a large scale observational study, like Markussen and Røed (2014)), or by exploiting additional truly random variation in practice regimes (as we do in our original analysis). Viewed this way, the results from this second robustness test serve a dual purpose. First, they substantiate our previous findings, and second, they illustrate that we are benefiting from the additional truly random variation in practice regimes induced by the experiment—a unique characteristic of the paper at hand.

The third robustness check is intended to test whether our results still apply when an alternative set of instruments is used. The instruments used so far were derived from a competing risks transition rate model: we used the (leave-out) mean covariate-adjusted transition propensities (into alternative activities) of all sick-listed workers exposed to a particular practice regime as instruments for actual participation in some activation requirements. The instruments used in this robustness test are simpler to construct. We simply estimate—separately for each of the three alternative treatments—the following linear probability model (one observation per sick-listed worker):

$$P_{Ai} = \mu_0 + \mu_1 X_i + \mu_2 X_j + \mu_C C + u_{A_{ij}} \tag{6}$$

where C is a vector of "center dummies", one for each practice regime. Note that the 32 elements of μ_C can be interpreted as covariate-adjusted indicators of a practice regime to use activity A. Therefore, μ_C provides a natural proxy for unobserved characteristics of different practice styles. Similarly to the rationale discussed in Section 5.1, it can be argued that these proxies for "practice regimes" provide legitimate instruments for the potentially endogenous program participation indicators. Table 7 below reports local average treatment effects estimated by using μ_C as instrument for P_{Ai} (see equation (2) for the structural model). It can immediately be seen that the estimates reported here correspond quite well with our main results (Table 4).

25

Table 6: Effects of participating in alternative treatment activities (LATE) evaluated at different points in time after randomization, using control group data only

	Dependent variable: Number of weeks in ...					
	I Regular employ- ment	II Self- sufficiency	III Sickness	IV Unemploy- ment	V Early retirement	VI Fleksjob
Panel A: *During 1st year after random assignment (N=2,304)*						
Traditional activation	1.266	-21.257	-11.582	19.868	0.110	3.680
	(24.452)	(24.599)	(26.406)	(15.529)	(3.636)	(3.885)
Paramedical care	-42.899	-50.258*	54.086*	6.216	-2.025	-1.363
	(28.897)	(29.070)	(31.206)	(18.352)	(4.296)	(4.591)
Graded return	18.991	8.493	-16.335	8.698	0.134	-2.754
	(12.080)	(12.152)	(13.044)	(7.671)	(1.796)	(1.919)
Panel B: *During 2nd year after random assignment (N=2,304)*						
Traditional activation	-34.025	-36.207	-13.958	27.092	5.141	10.010
	(28.503)	(29.665)	(22.464)	(22.195)	(11.047)	(10.604)
Paramedical care	-12.943	-39.734	35.473	-21.887	20.719	2.284
	(33.684)	(35.057)	(26.547)	(26.229)	(13.055)	(12.532)
Graded return	6.178	-0.098	-15.939	13.628	0.860	-4.076
	(14.080)	(14.655)	(11.097)	(10.964)	(5.457)	(5.239)
Panel C: *During 3rd year after random assignment (N=2,297)*						
Traditional activation	-19.581	-6.528	-19.304	-0.950	10.070	23.426
	(29.738)	(29.972)	(19.554)	(21.476)	(17.719)	(16.647)
Paramedical care	-5.076	-24.484	6.468	-4.116	38.658*	-5.539
	(34.357)	(34.628)	(22.592)	(24.811)	(20.472)	(19.233)
Graded return	11.914	12.720	-15.767	4.051	7.506	-3.900
	(14.783)	(14.899)	(9.721)	(10.676)	(8.808)	(8.275)

Notes: Second robustness test. The table shows the estimated effects of participating in alternative activation requirements (LATE). Each panel is based on six separate IV/2SLS regressions (second stage). See notes to Table A6 for a description of the outcome variables.

Included instruments: see Table A1 and Table A2 in the Appendix; week of intake dummies.

Excluded instruments: vector of local treatment strategy characteristics (PS_i). F test of excluded instruments (first stage): traditional activation: 7.14; paramedical care: 8.34; graded return: 9.24.

Standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Effects of participating in alternative treatment activities (LATE) evaluated at different points in time after randomization using center fixed effects as instruments for treatment participation (IV/2SLS)

| | Dependent variable: Number of weeks in ... | | | | | |
| | I | II | III | IV | V | VI |
	Regular employ- ment	Self- sufficiency	Sickness	Unemploy- ment	Early retirement	Fleksjob
Panel A: *During 1st year after random assignment (N=4,688)*						
Traditional activation	-3.174	-3.280	-1.178	2.628*	0.449	0.782*
	(2.334)	(2.354)	(2.510)	(1.464)	(0.409)	(0.471)
Paramedical care	-3.852**	-3.858**	2.917	-0.058	0.158	0.758**
	(1.747)	(1.762)	(1.879)	(1.095)	(0.306)	(0.353)
Graded return	14.388**	12.009**	-12.196*	-0.559	0.225	-2.177*
	(5.835)	(5.885)	(6.275)	(3.659)	(1.022)	(1.178)
Panel B: *During 2nd year after random assignment (N=4,688)*						
Traditional activation	-3.553	-1.708	-2.171	1.780	0.626	1.196
	(2.829)	(2.839)	(2.128)	(2.126)	(1.116)	(1.168)
Paramedical care	-5.799***	-6.485***	1.089	3.785**	1.049	1.405
	(2.117)	(2.125)	(1.592)	(1.591)	(0.835)	(0.875)
Graded return	18.750***	12.195*	-1.520	-14.042***	-0.040	-3.251
	(7.072)	(7.097)	(5.319)	(5.314)	(2.790)	(2.922)
Panel C: *During 3rd year after random assignment (N=4,680)*						
Traditional activation	-5.012*	-2.692	-0.934	-0.201	0.894	0.579
	(2.932)	(2.932)	(1.702)	(2.116)	(1.598)	(1.639)
Paramedical care	-1.216	-1.522	-1.432	1.897	1.976*	1.160
	(2.178)	(2.177)	(1.264)	(1.571)	(1.187)	(1.217)
Graded return	15.302**	7.587	3.395	-7.642	-2.345	-2.314
	(7.266)	(7.264)	(4.217)	(5.244)	(3.960)	(4.061)

Notes: Third robustness test. The table shows the estimated effects of participating in alternative activation requirements (LATE). Each panel is based on six separate IV/2SLS regressions (second stage). See notes to Table A6 for a description of the outcome variables.

Included instruments: see Table A1 and Table A2 in the Appendix; week of intake dummies.

Excluded instruments: center fixed effects; one center had to be excluded due to perfect multicollinearity in the linear probability model used to construct the instruments. F test of excluded instruments (first stage): traditional activation: 251.40; paramedical care: 540.22; graded return: 33.26.

Standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6   Conclusion

This paper has provided new and important evidence relating to the effects of a stronger activation strategy for sick-listed workers on their subsequent labour market outcomes. We have used a unique dataset from a large-scaled randomized experiment conducted in Danish job centers in 2009, linked to large administrative registers.

We first evaluated the *intention-to-treat* effect of participating in more activation requirements— whichever type—by simply comparing the average labour market outcomes of treated and non-treated ("difference in means"). Second, we have exploited exogenous variations in practice regimes, both between job centers and between treatment and control groups within a given job center, to compare the relative effectiveness of the alternative activities—intensified traditional activation, paramedical care, and graded return-to-work programs.

Our findings reveal first that the experimental intervention as a whole has been ineffective. Sick-listed workers initially assigned to the treatment group spent less time in regular employment and self-sufficiency (i.e., all forms of non-benefit receipt) compared to their peers in the control group who benefited from the standard activation strategy. Nevertheless, our results also show that a greater use of partial sick leaves is associated with an increase in regular employment and self-sufficiency, and a decrease in unemployment. Traditional activation and paramedical care, on the other hand, appear to have either no or even adverse impacts.

Taken together, our results thus suggest that graded return-to-work programs are the most effective intervention for improving sick-listed workers' subsequent labour outcomes. When intensified, they are associated with strong and lasting effects, for sick-listed workers who were employed before becoming sick and who do not suffer from mental disorders. Recent medical research also finds in favor of an increased emphasis on graded return-to-work (see e.g., Waddell and Burton (2006)).

References

Aakvik A., Heckman J., and Vytlacil E. (2005). Estimating Treatment Effects for Discrete Outcomes when Responses to Treatment Vary: An Application to Norwegian Vocational Rehabilitation Programs. *Journal of Econometrics*, 125(1-2): 15–51.

Aizer A. and Doyle J. (2015). Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges. *The Quarterly Journal of Economics*, 130(2): 759–803.

Andrén D. and Svensson M. (2012). Part-Time Sick Leave as a Treatment Method for Individuals with Musculoskeletal Disorders. *Journal of Occupational Rehabilitation*, 22(3): 418–426.

Angrist J. and Pischke J.S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Boll J., Hertz M., Rosholm M., and Svarer M. (2010). Evaluering af Aktive - Hurtigere Tilbage. Evaluering, Arbejdsmarkedsstyrelsen.

Card D., Kluve J., and Weber A. (2010). Active Labour Market Policy Evaluations: A Meta-Analysis. *Economic Journal*, 120(548): 452–477.

Dahl G., Kostøl A., and Mogstad M. (2014). Family Welfare Cultures. *The Quarterly Journal of Economics*, 129(4): 1711–1752.

Duggan M. (2005). Do New Prescription Drugs Pay for Themselves? The Case of Second-Generation Antipsychotics. *Journal of Health Economics*, 24(1): 1–31.

Engström P., Hägglund P., and Johansson P. (2015). Early Interventions and Disability Insurance: Experience from a Field Experiment. *Forthcoming in The Economic Journal.*

Everhardt T. and de Jong P. (2011). Return to Work After Long Term Sickness. *De Economist*, 159(3): 361–380.

Frölich M., Heshmati A., and Lechner M. (2004). A Microeconometric Evaluation of Rehabilitation of Long-term Sickness in Sweden. *Journal of Applied Econometrics*, 19(3): 375–396.

Hesselius P. (2007). Does Sickness Absence Increase the Risk of Unemployment? *The Journal of Socio-Economics*, 36(2): 288–310.

Heymann J., Rho H.J., Schmitt J., and Earle A. (2010). Ensuring a Healthy and Productive Workforce: Comparing the Generosity of Paid Sick Day and Sick Leave Policies in 22 Countries. *International Journal of Health Services*, 40(1): 1–22.

Høgelund J., Holm A., and Eplov L. (2012). The Effect of Part-time Sick Leave for Employees with Mental Disorders. *Journal of Mental Health Policy and Economics*, 15(4): 157–170.

Høgelund J., Holm A., and McIntosh J. (2010). Does Graded Return-to-work Improve Sick-listed Workers' Chance of Returning to Regular Working Hours? *Journal of Health Economics*, 29(1): 158–169.

Imbens G. and Angrist J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2): 467–75.

Kausto J., Miranda H., Martimo P., and Viikari-Juntura E. (2008). Partial Sick Leave: Review of Its Use, Effects and Feasibility in the Nordic Countries. *Scandinavian Journal of Work, Environment and Health*, 34(4): 239–249.

Kling J. (2006). Incarceration Length, Employment, and Earnings. *American Economic Review*, 96(3): 863–876.

Kools L. and Koning P. (2018). Graded Return-to-work as a Stepping Stone to Full Work Resumption. Mimeo.

Maestas N., Mullen K., and Strand A. (2013). Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt. *American Economic Review*, 103(5): 1797–1829.

Markussen S. (2012). The Individual Cost of Sick Leave. *Journal of Population Economics*, 25(4): 1287–1306.

Markussen S., Mykletun A., and Røed K. (2012). The Case for Presenteeism - Evidence from Norway's Sickness Insurance Program. *Journal of Public Economics*, 96(11): 959–972.

Markussen S. and Røed K. (2014). The Impacts of Vocational Rehabilitation. *Labour Economics*, 31: 1–13.

Nielsen M.B., Hansen J.V., Aust B., Tverborgvik T., Thomsen B., Bjorner J.B., Mortensen O.S., Rugulies R., Winzor G., Ørbæk P., Helverskov T., Kristensen N., and Poulsen O.M. (2014). A Multisite Randomized Controlled Trial on Time to Self-support among Sickness Absence Beneficiaries. The Danish National Return-to-work Programme. *European Journal of Public Health*.

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

Røed K. (2012). Active Social Insurance. *IZA JOurnal of Labor Policy*, 1(8): 1–22.

Van Oostrom S., Driessen M., de Vet H., Franche R., Schonstein E., Loisel P., Mechelen W., and Anema J. (2009). *Workplace Interventions for Preventing Work Disability (Review).* The Cochrane Library. John Wiley & Sons, Ltd.

Viikari-Juntura E., Kausto J., Shiri R., Kaila-Kangas L., Takala E., Karppinen J., Miranda H., Luukkonen R., and Martimo P. (2012). Return to Work After Early Part-time Sick Leave due to Musculoskeletal Disorders: a Randomized Controlled Trial. *Scandinavian Journal of Work, Environment and Health*, 38(2): 134–143.

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

Appendix

Table A1: Pre-treatment characteristics by treatment status (N=4,728)

Variable	Control group		Treatment group		Difference
	Mean	Std. dev.	Mean	Std. dev.	
Panel A: *Socio-demographic characteristics*					
< 30 years old	0.150	(0.357)	0.137	(0.343)	-0.014
30-39 years old	0.234	(0.423)	0.218	(0.413)	-0.015
40-49 years old	0.273	(0.446)	0.286	(0.452)	0.013
> 49 years old	0.343	(0.475)	0.359	(0.480)	0.016
Male	0.433	(0.496)	0.457	(0.498)	0.024
Married	0.583	(0.493)	0.570	(0.495)	-0.013
Danish origin	0.878	(0.327)	0.876	(0.329)	-0.002
Western origin	0.045	(0.207)	0.050	(0.217)	0.005
Non-western origin	0.077	(0.267)	0.074	(0.262)	-0.003
Panel B: *Labor market history*					
Sick-listed from regular empl.	0.777	(0.416)	0.743	(0.437)	-0.034***
Sick-listed from UI-benefits	0.160	(0.366)	0.194	(0.395)	0.034***
Sick-listed from self-empl.	0.063	(0.243)	0.063	(0.242)	-0.000
Elapsed sickness duration^a	8.902	(6.235)	8.601	(6.082)	-0.300*
Time spent on sickness-ben.^{b}	0.212	(0.143)	0.208	(0.145)	-0.003
Time spent on sickness-ben.^{c}	0.080	(0.176)	0.080	(0.175)	-0.000
Time spent on sickness-ben.^{d}	0.069	(0.163)	0.063	(0.156)	-0.006
Degree of pub. inc. support$^{b\,e}$	0.311	(0.251)	0.317	(0.259)	0.006
Degree of pub. inc. support$^{c\,e}$	0.216	(0.314)	0.213	(0.314)	-0.003
Degree of pub. inc. support$^{d\,e}$	0.240	(0.338)	0.232	(0.332)	-0.007
Panel C: *Diagnoses*					
Mental health disorder	0.334	(0.472)	0.319	(0.466)	-0.015
Non-mental disorder	0.666	(0.472)	0.681	(0.466)	0.015
Number of observations	2,304		2,424		

^{a} at start of experiment (in weeks) ^{b} in year before sickness ^{c} in second last year before sickness ^{d} in third last year before sickness ^{e} any kind of public income support.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Municipality-level controls used both in competing risks model and in main outcome equation (local treatment strategies approach)

Variable	Description
Panel A: *Socio-demographic characteristics*	
avage	Average age
fertil	Total fertility rate
lifeexp	Life expectancy for new born babies
danes	Persons of Danish origin in % of the population
lowedu	Percentage of working-age population with no more than ten years of schooling
migra1	External net migration in % of the population
migra2	Internal net migration in % of the population
crime	Reported criminal offences per capita
Panel B: *Local labour market conditions*	
unemp	Full-time unemployed persons in % of the labour force
sbens1	Persons receiving sickness benefits (w/ job) in % of the working-age population
sbens2	Persons receiving sickness benefits (w/o job) in % of the working-age population
lfpart	Economic activity rate (labour force participation)
outlf	Percentage of working-age population outside the labour force
prisec	Persons employed in primary sector in % of the population
vacs1	Number of open vacancies per unemployed
vacs2	Number of newly opened vacancies per unemployed

 Notes: Data on all variables (except for *vacs1* and *vacs2*) stem from Statistics Denmark, Denmark's national statistics institute.

Table A3: Unconditional intention-to-treat effects at different points in time after randomization

	Time after randomization					
	One year		Two years		Three years	
Number of weeks in...	ITT	Mean in control group	ITT	Mean in control group	ITT	Mean in control group
Regular employment	-1.390***	16.474	-1.390***	20.865	-0.842	20.675
	(0.505)	(0.516)	(0.505)	(0.516)	(0.607)	(0.433)
Self-sufficiency	-1.395***	19.828	-1.891***	25.774	-1.065*	25.705
	(0.516)	(0.370)	(0.599)	(0.426)	(0.614)	(0.435)
Sickness	0.416	19.812	0.265	8.141	-0.287	5.197
	(0.498)	(0.356)	(0.394)	(0.275)	(0.311)	(0.223)
Unemployment	0.409	4.779	0.586	7.545	0.436	7.366
	(0.286)	(0.199	(0.402)	(0.279)	(0.408)	(0.285)
Early retirement	0.599***	0.265	0.731***	1.244	0.862***	1.705
	(0.112)	(0.059)	(0.228)	(0.136)	(0.264)	(0.168)
Disability	-0.064***	0.503	0.034	1.768	0.328	2.871
	(0.107)	(0.081)	(0.229)	(0.162)	(0.313)	(0.217)
Fleksjob	0.222***	1.739	0.620***	2.527	0.665*	3.609
	(0.240)	(0.167)	(0.293)	(0.197)	(0.354)	(0.241)
N	5,652		5,652		5,644	

Notes: Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Conditional intention-to-treat effects at different points in time after randomization, by labour market status before sickness

| Number of weeks in... | Time after randomization | | |
| | One year | Two years | Three years |
	ITT	ITT	ITT
Panel A—status before sickness: employed			
Regular employment	-0.975*	-1.375**	-0.400
	(0.554)	(0.651)	(0.675)
Self-sufficiency	-1.137***	-1.407**	-0.867
	(0.553)	(0.640)	(0.663)
Sickness	0.039	0.180	-0.179
	(0.519)	(0.424	(0.333)
Unemployment	0.108	0.203	0.131
	(0.276)	(0.402)	(0.417)
Early retirement	0.647***	0.847***	0.964***
	(0.129	(0.254)	(0.289)
Disability	-0.053	-0.068	0.150
	(0.125)	(0.260)	(0.349)
Fleksjob	0.385	0.751**	0.749*
	(0.286)	(0.342)	(0.399)
N	4,412	4,412	4,407
Panel B—status before sickness: unemployed			
Regular employment	-1.067	-1.522	-1.08
	(0.847)	(1.145)	(1.238)
Self-sufficiency	-1.009	-2.098*	-1.628
	(0.897)	(1.262)	(1.339)
Sickness	0.461	-0.116	-0.763*
	(1.295)	(1.126)	(0.951)
Unemployment	-0.070	0.511***	0.553
	(0.994)	(1.265)	(1.262)
Early retirement	0.581**	0.797	1.014
	(0.279)	(0.561)	(0.658)
Disability	-0.051	0.462	0.960
	(0.277)	(0.604)	(0.861)
Fleksjob	0.138	0.676	0.654
	(0.110)	(0.573)	(0.850)
N	894	894	891

Notes: Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Conditional intention-to-treat effects at different points in time after randomization, by diagnosis

| | Time after randomization | | |
| | One year | Two years | Three years |
Number of weeks in...	ITT	ITT	ITT
Panel A—status before sickness: non-mental disorders			
Regular employment	-1.791***	-2.231***	-1.197***
	(0.609)	(0.706)	(0.723)
Self-sufficiency	-1.691***	-1.971***	-1.154
	(0.624)	(0.713)	(0.734)
Sickness	0.871	0.442	-0.161
	(0.594)	(0.459)	(0.368)
Unemployment	0.300	0.483	0.511
	(0.324)	(0.452)	(0.455)
Early retirement	0.608***	0.784***	0.929***
	(0.133)	(0.283)	(0.332)
Disability	-0.064	0.071	0.435
	(0.127)	(0.276)	(0.376)
Fleksjob	0.278	0.546	0.616
	(0.278)	(0.373)	(0.438)
N	3,979	3,979	3,976
Panel B—status before sickness: mental disorders			
Regular employment	-0.354	-0.791	0.081
	(0.834)	(1.012)	(1.049)
Self-sufficiency	-0.697***	-1.821*	-0.916
	(0.834)	(1.012)	(1.049)
Sickness	-0.579	-0.084	-0.533
	(0.864)	(0.747)	(0.580)
Unemployment	0.663	0.830	0.316
	(0.568)	(0.783)	(0.804)
Early retirement	0.562***	0.555	0.618
	(0.203)	(0.358)	(0.388)
Disability	-0.036	-0.006	0.152
	(0.201)	(0.405)	(0.551)
Fleksjob	0.087	0.784*	0.792
	(0.319)	(0.438)	(0.585)
N	1,673	1,673	1,668

Notes: Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: ITT effects of marginal changes in local treatment strategies at different points in time after randomization (OLS, reduced form estimates)

	Dependent variable: Number of weeks in ...					
	I Regular employ-ment	II Self-sufficiency	III Sickness	IV Unemploy-ment	V Early retirement	VI Fleksjob
Panel A: *During 1st year after random assignment (N=4,728)*						
$\varphi_{traditional\,activation}$	-2.476* (1.381)	-2.133 (1.401)	-0.444 (1.444)	1.408 (0.859)	0.280 (0.239)	0.768*** (0.268)
$\varphi_{paramedical\,care}$	-2.937** (1.161)	-2.896** (1.178)	2.073* (1.214)	-0.368 (0.722)	0.119 (0.201)	0.734*** (0.225)
$\varphi_{graded\,return}$	4.227*** (1.578)	3.245** (1.602)	-3.538** (1.651)	0.251 (0.981)	0.125 (0.273)	-0.862*** (0.306)
Panel B: *During 2nd year after random assignment (N=4,728)*						
$\varphi_{traditional\,activation}$	-3.398** (1.656)	-2.610 (1.685)	-0.612 (1.248)	2.077* (1.207)	0.378 (0.655)	1.374** (0.685)
$\varphi_{paramedical\,care}$	-3.345** (1.392)	-4.055*** (1.417)	1.006 (1.049)	1.304 (1.015)	0.628 (0.550)	1.599*** (0.576)
$\varphi_{graded\,return}$	3.674* (1.893)	2.202 (1.927)	-1.446 (1.426)	-2.457* (1.380)	0.622 (0.748)	-1.382* (0.783)
Panel C: *During 3rd year after random assignment (N=4,720)*						
$\varphi_{traditional\,activation}$	-3.199* (1.707)	-2.065 (1.723)	-0.458 (0.985)	0.515 (1.231)	0.616 (0.935)	1.044 (0.958)
$\varphi_{paramedical\,care}$	-0.982 (1.436)	-1.552 (1.450)	-0.594 (0.829)	0.652 (1.035)	0.860 (0.787)	1.531* (0.806)
$\varphi_{graded\,return}$	3.689* (1.956)	1.906 (1.974)	-0.199 (1.128)	-1.664 (1.410)	0.794 (1.072)	-1.355 (1.098)

Notes: The table shows ITT effects of marginal changes in practice regimes (reduced form estimates). Each panel is based on six separate OLS regressions of the number of weeks in regular (i.e., wage) employment/self-sufficiency/sickness/unemployment/early retirement/fleksjob on the vector of practice regimes and additional controls (individual and municipality-level socio-economic characteristics; week of intake dummies). Following Markussen and Røed (2014: 19), we have normalized the vectors of practice regimes. Consequently, these parameter estimates can be interpreted as the expected change in the outcome variable "resulting from a movement from the practice regime giving lowest priority to the strategy under consideration to the one giving it highest priority". Standard errors in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Effects of participating in alternative activation requirements (LATE) evaluated at different points in time after randomisation, by labour market status before sickness

	I Regular employment	II Self-sufficiency	III Sickness	IV Unemployment	V Early retirement	VI Fleksjob
	Dependent variable: Number of weeks in ...					

(Dependent variable: Number of weeks in ...)

	I Regular employment	II Self-sufficiency	III Sickness	IV Unemployment	V Early retirement	VI Fleksjob
Panel A: *Labor market status before sickness: Regular employment*						
Panel A1: *During 1st year after random assignment (N=3,593)*						
Traditional activation	-4.896	-3.413	-1.712	3.438*	0.552	1.375
	(3.932)	(3.741)	(3.964)	(1.852)	(0.526)	(0.883)
Paramedical care	-9.990**	-9.896**	8.836**	-2.060	0.318	2.449***
	(4.138)	(3.936)	(4.171)	(1.948)	(0.554)	(0.930)
Graded return	36.255**	31.223**	-32.384**	3.446	0.650	-7.333**
	(14.543)	(13.833)	(14.660)	(6.848)	(1.945)	(3.267)
Panel A2: *During 2nd year after random assignment (N=3,593)*						
Traditional activation	-4.193	-0.287	-2.460	2.568	-0.242	2.166
	(4.103)	(3.987)	(2.901)	(2.610)	(1.535)	(1.833)
Paramedical care	-9.454**	-10.832***	4.597	1.749	0.706	4.447**
	(4.318)	(4.196)	(3.053)	(2.747)	(1.615)	(1.929)
Graded return	26.203*	20.595	-13.074	-9.846	4.192	-12.261*
	(15.174)	(14.746)	(10.728)	(9.654)	(5.675)	(6.780)
Panel A3: *During 3rd year after random assignment (N=3,588)*						
Traditional activation	-7.247*	-3.890	-0.365	1.337	-0.603	1.912
	(4.352)	(4.318)	(2.181)	(2.667)	(2.160)	(2.499)
Paramedical care	-4.892	-7.580*	0.389	-1.388	1.236	5.278**
	(4.605)	(4.569)	(2.308)	(2.822)	(2.286)	(2.645)
Graded return	28.269*	26.934*	-4.043	-1.991	4.474	-16.558*
	(16.264)	(16.134)	(8.152)	(9.967)	(8.071)	(9.340)
Panel B: *Labor market status before sickness: Unemployment*						
Panel B1: *During 1st year after random assignment (N=838)*						
Traditional activation	-0.681	-0.329	1.857	1.526	-0.580	0.087
	(3.288)	(3.535)	(5.200)	(4.220)	(1.210)	(0.590)
Paramedical care	-1.579	-1.604	-0.517	-2.321	0.358	0.762
	(3.286)	(3.533)	(5.196)	(4.217)	(1.210)	(0.589)
Graded return	0.464	-16.020	44.974	26.351	-18.111	-8.015
	(48.671)	(52.337)	(76.974)	(62.464)	(17.917)	(8.730)
Panel B2: *During 2nd year after random assignment (N=838)*						
Traditional activation	-6.925	-4.427	4.205	2.024	-0.856	-0.124
	(4.761)	(5.515)	(4.841)	(5.601)	(2.566)	(2.625)
Paramedical care	2.498	0.874	-3.581	0.057	0.196	4.346*
	(4.758)	(5.512)	(4.838)	(5.597)	(2.564)	(2.623)
Graded return	-44.981	-71.754	19.302	46.755	5.994	-22.836
	(70.473)	(81.641)	(71.669)	(82.910)	(37.981)	(38.854)
Panel B3: *During 3rd year after random assignment (N=835)*						
Traditional activation	-5.403	-2.365	4.992	-6.319	2.117	1.954
	(5.213)	(5.454)	(5.258)	(5.641)	(3.976)	(3.468)
Paramedical care	4.302	-0.143	-7.702	9.036	-1.722	2.705
	(5.189)	(5.429)	(5.234)	(5.616)	(3.958)	(3.452)
Graded return	-59.668	-58.421	90.707	-63.715	42.490	-1.651
	(76.900)	(80.452)	(77.554)	(83.215)	(58.647)	(51.155)

Notes: Standard errors in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Effects of participating in alternative activation requirements (LATE) evaluated at different points in time after randomisation, by diagnosis

	I Regular employment	II Self-sufficiency	III Sickness	IV Unemployment	V Early retirement	VI Fleksjob
	Dependent variable: Number of weeks in ...					

Panel A: *Sickness type: Non-mental sickness*

Panel A1: *During 1st year after random assignment (N=3,184)*

	I Regular employment	II Self-sufficiency	III Sickness	IV Unemployment	V Early retirement	VI Fleksjob
Traditional activation	-4.275	-4.743	1.588	2.321	0.791	1.658
	(3.802)	(3.776)	(3.881)	(1.971)	(0.617)	(1.036)
Paramedical care	-6.560*	-6.144*	4.777	-1.789	0.005	2.129**
	(3.501)	(3.477)	(3.574)	(1.815)	(0.568)	(0.954)
Graded return	30.801**	28.375*	-25.392	0.511	1.267	-10.177**
	(15.312)	(15.208)	(15.631)	(7.936)	(2.483)	(4.170)

Panel A2: *During 2nd year after random assignment (N=3,184)*

Traditional activation	-6.049	-6.025	1.812	2.211	0.945	1.949
	(4.228)	(4.241)	(3.277)	(3.071)	(1.729)	(1.994)
Paramedical care	-7.342*	-7.272*	3.759	2.452	0.552	3.205*
	(3.893)	(3.905)	(3.018)	(2.828)	(1.593)	(1.836)
Graded return	27.197	23.607	-19.251	-19.809	5.410	-12.903
	(17.028)	(17.081)	(13.198)	(12.369)	(6.965)	(8.029)

Panel A3: *During 3rd year after random assignment (N=3,181)*

Traditional activation	-7.602	-5.802	0.478	1.556	0.897	2.039
	(4.647)	(4.435)	(2.573)	(3.015)	(2.320)	(2.470)
Paramedical care	-3.882	-3.129	1.729	1.459	2.473	2.038
	(4.273)	(4.078)	(2.366)	(2.772)	(2.133)	(2.271)
Graded return	35.566*	26.155	-12.268	-16.986	0.240	-9.422
	(18.597)	(17.752)	(10.297)	(12.066)	(9.283)	(9.886)

Panel B: *Sickness type: Mental sickness*

Panel B1: *During 1st year after random assignment (N=1,544)*

Traditional activation	-1.284	5.830	2.141	-10.680	1.170	2.952
	(21.214)	(21.007)	(60.340)	(72.118)	(8.751)	(12.660)
Paramedical care	-8.616	-0.177	28.233	-31.032	4.343	6.200
	(42.464)	(42.048)	(120.781)	(144.360)	(17.516)	(25.341)
Graded return	36.272	-16.177	-139.830	165.143	-19.951	-29.125
	(219.939)	(217.785)	(625.578)	(747.704)	(90.724)	(131.251)

Panel B2: *During 2nd year after random assignment (N=1,544)*

Traditional activation	3.549	18.189	-38.318	21.028	-0.846	11.731
	(26.561)	(63.758)	(129.939)	(58.765)	(9.985)	(49.175)
Paramedical care	2.487	18.562	-57.365	27.807	4.719	24.052
	(53.166)	(127.623)	(260.097)	(117.628)	(19.988)	(98.432)
Graded return	-25.376	-132.725	303.553	-137.756	-14.903	-115.386
	(275.368)	(661.013)	(1347.146)	(609.241)	(103.524)	(509.820)

Panel B3: *During 3rd year after random assignment (N=1,539)*

Traditional activation	15.504	24.716	-37.859	-28.692	-22.179	30.706
	(155.994)	(197.655)	(324.129)	(270.304)	(186.662)	(289.803)
Paramedical care	32.379	38.549	-76.287	-59.552	-40.507	68.841
	(318.409)	(403.438)	(661.598)	(551.730)	(381.007)	(591.543)
Graded return	-172.082	-225.795	391.907	319.714	221.979	-348.506
	(1,668.178)	(2,113.648)	(3,466.182)	(2,890.555)	(1,996.136)	(3,099.160)

Notes: Standard errors in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$