I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13971

Do Disability Benefits Hinder Work Resumption after Recovery?

Pierre Koning
Paul Muller
Roger Prudon

DECEMBER 2020

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13971

Do Disability Benefits Hinder Work Resumption after Recovery?

Pierre Koning
Vrije Universiteit Amsterdam, Tinbergen Institute, Leiden University and IZA

Roger Prudon
Vrije Universiteit Amsterdam and Tinbergen Institute

Paul Muller
Vrije Universiteit Amsterdam and Tinbergen Institute

DECEMBER 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Do Disability Benefits Hinder Work Resumption after Recovery?*

While a large share of Disability Insurance recipients are expected to recover, outflow rates from temporary disability schemes are typically negligible. We estimate the disincentive effects of disability benefits on the response to a (mental) health improvement using administrative data on all Dutch disability benefit applicants. We compare those below the DI eligibility threshold with those above and find that disincentives significantly reduce work resumption after health improves. Approximately half of the response to recovery is offset by benefits. Structural labor supply model estimates suggest disincentive effects are substantially larger when the workers earnings capacity is fully restored.

JEL Classification: J08, I1, J22

Keywords: disability insurance, mental health, labor supply, health shocks

Corresponding author:
Pierre Koning
VU University Amsterdam
Department of Economics, 7A-27
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands
E-mail: p.w.c.koning@vu.nl

* We thank Owen O'Donnell, Maarten Lindeboom and seminar participants at the VU Amsterdam, University of Amsterdam, Gothenburg University, EEA Congress and EALE conference for useful comments. The authors gratefully acknowledge the financial support from Instituut Gak for this research project, as well as UWV for providing access to their data on DI applications. Results in this paper are based on non-public microdata from Statistics Netherlands (CBS). In principle, part of these microdata are accessible for statistical and scientific research. For further information: microdata@cbs.nl.

1 Introduction

One of the largest social insurance schemes in developed economies is Disability Insurance (DI) (OECD, 2019). In the Netherlands for example, approximately 9% of the working population received disability benefits in 2017 whereas respectively 4% and 4.5% of the working population received unemployment benefits and social assistance benefits (CBS, 2019). Total spending on disability benefits amounted to 1.5% of GDP making it the most sizable social insurance scheme in terms of expenditure. In most OECD countries, a large share of these expenditures is intended for temporary disabilities (Pettersson-Lidbom & Thoursie, 2013), of which mental health problems are one of the main causes.[1] Even though a large share of those with mental health problems are expected to recover (Korpi, 2001), their work resumption rate is typically low (Claussen, Bjørndal, & Hjort, 1993; Autor & Duggan, 2006).

Despite the stark discrepancy between expected and realized outflow from disability, the vast majority of research on disability insurance has focused on policies targeting inflow or the use of schemes.[2] As one of the few studies that considers DI outflow, we investigate the role of mental health recovery in the return to employment for DI recipients. Comparing DI applicants with disability degrees below and above the insurance cutoff, we estimate the effect of receiving benefits on the labor supply response to a positive health shock that is measured by mental health recovery. We assess whether disability benefits create disincentives for returning to work once health improves. Considering all DI applicants in the Netherlands since 2006, we show that health improvements indeed coincide with an increase in labor supply. At the same time, disincentives from receiving benefits do matter, as awarded applicants with partial disability benefits show weaker labor supply responses than those without benefits.

The main reason for the lack of research on the interplay of disability benefits, health recovery and labor supply is the absence of reliable data on the dynamics of individuals' health. We address this problem by linking three sources of Dutch admin-

[1] For example, Maestas (2019) emphasizes the substantial increase in prevalence of mental health problems among DI recipients in the US and points towards partial and temporary DI benefits as promising reforms for the DI system.

[2] By now there is a vast literature on moral hazard and targeting of DI benefits. Specifically, higher benefits lead to higher inflow rates (Gruber, 2000; Borghans, Gielen, & Luttmer, 2014) and more stringent selection criteria reduce inflow (Staubli, 2011; Godard, Koning, & Lindeboom, 2019).

istrative data covering the entire population of DI applicants since 2006 (over 600,000 individuals): (i) DI application records (including detailed assessment outcomes), (ii) monthly administrative records on employment and social insurance recipients and (iii) administrative records describing mental health treatments. Using the DI application data, we are able to identify applicants above and below the disability benefits cutoff in terms of degree of disability. As these applicants are at the lower end of the disability severity distribution, they are typically expected to recover at some point. For those with mental health problems, we obtain a plausible measurement of significant health improvement by considering the end date of a mental health treatment. While certainly not being indicative of full recovery, we interpret the end of a mental health treatment trajectory as a substantial improvement in health. We compare labor supply responses around this end date for those that receive disability benefits with those that do not, yielding estimates of the employment disincentive effects of benefit receipt.

Our approach essentially constitutes a Difference-in-Differences (DiD) estimator that compares applicants with and without disability benefits, before and after recovery. Using employment histories we show that pre-trends for the two groups are parallel, at least until closely before mental health treatment ends. While one may worry that the end of mental health treatment is not a perfect proxy for recovery, our DiD estimator only requires the assumption that it proxies recovery equally well in the two groups. A similar argument holds for the issue of reverse causality: in some cases it may be employment that *causes* health improvement.[3] Again, any biasing effects are mitigated as long as reverse causality is equally strong in both groups. Using detailed health data, we show that the probability of recovery is very similar for both groups.

When using this design, it is important to control for the fact that individuals with DI benefits have a larger assessed loss of earnings capacity than those without DI benefits. Accordingly, the response to recovery might depend on the initial level of health. We therefore estimate this difference-in-health-effect by applying a DiD estimator on two adjacent groups of disability degree categories with equal DI benefit statuses (either awarded or rejected). Based on this information, we obtain an adjusted estimate of the DI benefit effect, namely by subtracting the difference-in-health-effect

[3]The reverse causal impact of employment on health has been assessed by using employment shocks such as mass lay-offs (Browning, Moller Dano, & Heinesen, 2006; Sullivan & Von Wachter, 2009) and firm closures (Morris & Cook, 1991; Kuhn, Lalive, & Zweimüller, 2009; Schmitz, 2011; Salm, 2009). The results are however ambiguous and inconclusive.

from the original DiD estimate.[4]

We observe that around the time of recovery the employment rates start to diverge, as those without disability benefits start working at a higher rate than those with disability benefits. The disincentives for work resumption are substantial, amounting to a negative impact of disability benefits of 2 percentage-points on employment, relative to baseline employment of around 30%. We interpret this as a large impact, since the pre-recovery difference between the groups is small and our proxy measures only partial recovery. Furthermore, the employment effect of recovery on individuals without DI benefits is 4 percentage-points, indicating that benefits absorb approximately half of this effect. Using a second proxy for health improvement that is based on significant drops in healthcare expenditures, we find similar results that extend to physical health improvements. Our findings are also robust against a series of alternative specifications, including imposing different 'donuts' around recovery to deal with imperfect measurement of the exact timing of recovery.

For a broader perspective on benefit disincentive effects, we next benchmark our estimates against predictions from a structural labor supply model. We estimate structural model parameters using information from pre-disability labor supply and the assessed remaining earnings capacity of workers. With this information, we validate the model by comparing predicted labor supply after a disability shock with the observed labor supply after a DI application. With the resulting calibrated model, we simulate labor supply responses to health recovery. Defining full recovery as a situation where earnings capacity *and* disutility from working return to their pre-application level, we find benefit disincentive effects of disability benefits equal to around 15 percentage-points. This estimate should be considered as an upper bound of the disincentive effect. Knowing that the end of the mental health treatment does not necessarily coincide with full recovery in all cases, it is not surprising to see considerably smaller reduced form effects.

Our findings contribute to the literature on the effects of financial incentives of DI schemes on employment. Disincentive effects of disability benefits on employment have been shown to exist in various settings. DI beneficiaries are less likely to be employed

[4]This method is comparable to a triple-difference estimator, in which we analyse whether the difference in the response to recovery differs if we analyse two groups with the same DI benefit status or two groups with a different DI benefit status.

than those whose DI application has been rejected (Bound, 1989; Chen & Van der Klaauw, 2008; Von Wachter, Song, & Manchester, 2011; Maestas, Mullen, & Strand, 2013; French & Song, 2014) and work resumption rates of DI beneficiaries also depend on financial incentives (Weathers & Hemmeter, 2011; Campolieti & Riddell, 2012; Kostol & Mogstad, 2014; Koning & van Sonsbeek, 2017). In a similar vein, DI beneficiaries whose benefits are terminated have high work resumption rates. In this respect, Moore (2015) exploits the exclusion of drug and alcohol addictions as qualifying conditions for the disability insurance system in the US, whereas Garcia Mandico, García Gómez, Gielen, and O'Donnell (2018) evaluate a reassessment of disability benefit claimants in the Netherlands. Both studies find increases in employment of approximately 20 percentage-points for those whose benefits were terminated. As reassessments in these studies, conducted some time after the initial assessment, coincide with tighter eligibility criteria, however, it is not possible to disentangle the effect from potential health recovery from the effect from tightened eligibility criteria.

The second relevant strand of related literature examines the relationship between health and labor supply. This literature either studies the effect of negative health shocks on labor supply or the effect of employment on health outcomes. Using self-assessed health (García Gómez & López Nicolás, 2006; Lindelow & Wagstaff, 2005), road injuries (Dano, 2005) and acute unscheduled hospitalizations (García Gómez, Van Kippersluis, O'Donnell, & Van Doorslaer, 2013; Lindeboom, Llena-Nozal, & van der Klaauw, 2016) a causal relationship has been established between negative health shocks and labor supply. These studies find that employment rates drop by 5 to 7 percentage-points after a negative health shock. The causal impact of health *improvements*, as considered in this paper, has not been examined so far due to absence of reliable data on positive health shocks. We thus add to this literature by incorporating disability insurance in the interplay between health and labor and by examining positive rather than negative health shocks.

The remainder paper is organized as follows. Section 2 illustrates the institutional background of the Dutch disability insurance system and section 3 gives a detailed description of the data. Section 4 provides a description of the Difference-in-Difference estimator and provides estimation results. Section 5 presents the structural labor supply model and section 6 concludes.

2 Disability insurance in the Netherlands

This section highlights the key characteristics of DI in the Netherlands, and the (dis)incentives it creates for applicants and those receiving benefits. The first subsection discusses the general historical background whereas the subsequent subsections explain the application process and the DI benefit system. For a more detailed description, we refer to Koning and Lindeboom (2015).

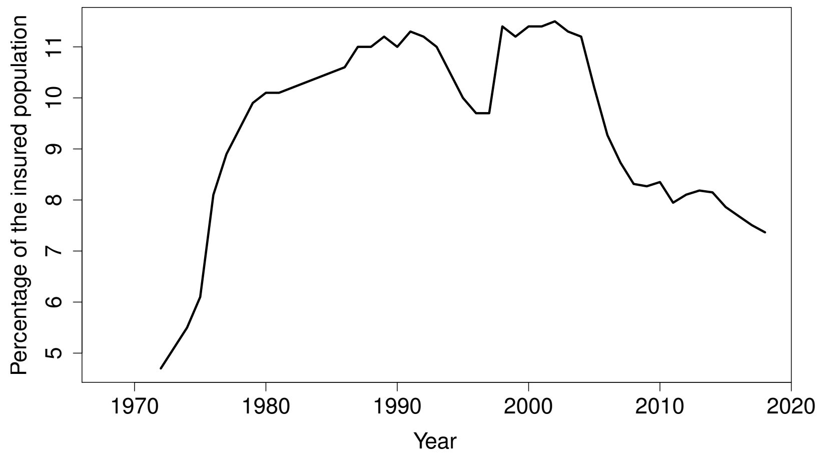
2.1 Historical background

The DI system is managed by the employee insurance agency (UWV). The proportion of people receiving disability benefits grew from approximately 4% of the working population to 12% of the working population between its implementation in the 1960's and the 1980's. It remained at approximately 12% until the beginning of the 2000's (Figure 1). From 1996 onward, several policy reforms were implemented in order to reduce the number of benefit recipients. One of these reforms (discussed in more detail below) was to replace the old disability benefit system (WAO) with a new disability benefit system (WIA). The new system comprises of a benefit system for partially and temporarily disabled individuals (WGA) and a system for fully and permanently disabled individuals (IVA). As shown in Figure 2, the successive reforms substantially reduced the inflow into DI, resulting in a decrease in the total amount of benefit recipients. The outflow rates however remained at approximately the same level since the 1980's, with the outflow rate of the system for temporarily disabled individuals even lower than the outflow rate before the reforms.

The reduction in inflow resulted from a series of policy changes (Koning, 2019). In 2002 the Gatekeeper Protocol ("Wet verbetering Poortwachter") was implemented. The protocol increased the responsibility of employer and employee in the waiting period leading up to the application. It specifies actions to be taken by both parties to accommodate the reintegration of the sick employee. The implementation of the Gatekeeper Protocol is estimated to have reduced inflow into disability insurance by 30 to 40% (Koning & Lindeboom, 2015; Van Sonsbeek & Gradus, 2012).

In 2006 several policy changes were implemented. First, eligibility criteria where tightened by widening the range of "reference" jobs, leading to lower disability benefits and higher rejection rates. Second, the waiting period (the period during which

Figure 1: DI recipients as a percentage of insured population in the Netherlands



Source: UWV (2012, 2018)

the employer continues wage payments to sick employees before DI application) was extended from one to two years. Lastly, the law on employment and income according to employment ability ("WIA") was implemented. This law increased the minimal required level of disability from 15 to 35%, and increased the incentive for DI benefit recipients to use their remaining earnings capacity. These policy changes further reduced the inflow into the disability insurance system (Koning & van Sonsbeek, 2017).

2.2 The disability insurance system

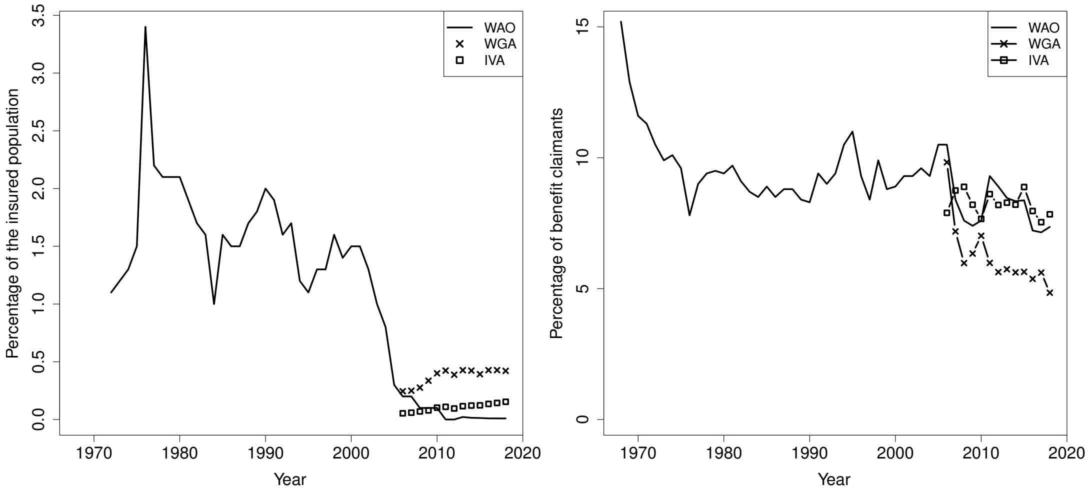
Our data covers all disability insurance applicants in the Netherlands between 2006 and 2017. In this period no major reforms were implemented. A DI application can be filed after two years of sickness.[5] Based on the assessed remaining earnings capacity and the pre-application earnings, the disability degree is determined as follows:

$$\text{Degree of disability} = (1 - \frac{\text{Remaining earnings capacity}}{\text{Pre-application earnings}}) * 100\% \tag{1}$$

Depending on the disability degree, individuals are assigned to one of the following disability degree brackets: 0-35%, 35-45%, 45-55%, 55-65%, 65-80% and 80-100%.

[5] The disability insurance process from the start of the illness until the actual application varies by type of employment contract. This process is described in detail in Online Appendix B.1.

Figure 2: Inflow and outflow rates of disability benefit programs (UWV (2012, 2018))



(a) DI inflow

(b) DI outflow

Disability benefits are awarded if the assessed degree of disability exceeds 35%. Benefits are based on the midpoint of the assigned disability interval. As an illustration, consider an individual with pre-application earnings of €3,000 per month. If remaining work capacity is 16 hours (i.e., 40% of full-time work) at a wage of €2,000 per month on a full-time basis, the remaining earnings capacity is set at €800. The resulting degree of disability is then 73.3%[6], implying that the relevant disability degree bracket is 65-80%.

If awarded benefits, benefit conditions differ between the so called "wage related period" and the "wage continuation period" (UWV, 2019b). The wage related period applies to anyone who worked at least 26 weeks within the 36 weeks prior to falling ill. The benefits amount to 70% of the difference between pre-application earnings and current earnings. The duration of the wage related period is roughly one month for every year worked since the age of 18.[7] When the wage related period ends, the wage continuation period starts. From then on, the benefits level depends on the utilization of the remaining earning capacity. If someone uses at least 50% of their remaining earning capacity, the amount of benefits remains linked to the pre-application earnings.

If earnings are equal to or exceed the assessed remaining earnings capacity, benefits

[6] $(1 - (800/3000)) * 100\% = 73.3\%$

[7] The setup of the disability system is slightly different for individuals with an assessed degree of disability above 80% whose disability is deemed to be permanent. They receive disability benefits amounting to 75% of their pre-application earnings and no re-assessments are performed

amount to 70% of the difference between pre-application earnings and currently earned earnings. If earnings are between 50% and 100% of the remaining earnings capacity, benefits amount to 70% of the difference between the pre-application earnings and the remaining earnings capacity. As stated earlier, wage earnings less than 50% of the remaining earning capacity imply that (partial) benefits are no longer linked to the pre-application earnings; they then equal 70% of the minimum wage, multiplied by the individuals' degree of disability. Individuals in the wage continuation period are not eligible for additional unemployment benefits, but might be able to apply for social assistance (approximately 70% of the minimum wage).

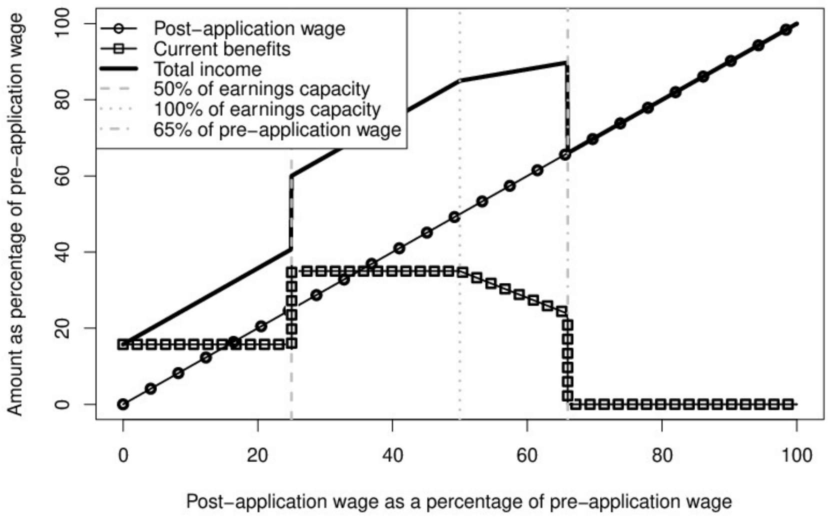
Finally, disability benefits can be terminated if the remaining earnings capacity is deemed to have increased over 65% of previous earnings. This can be due to increases in actual wages or medical reassessments. Reassessments can be requested by UWV (because of earnings changes), the former employer or by the recipient. Reassessments due to suspected changes in health are scarce compared to those due to changes in wage earnings (UWV, 2019a).

2.3 Work incentives

The DI system creates (dis)incentives to work that may differ in their impact before and after recovery. To illustrate this, we show an example in which the remaining earnings capacity is 50% of pre-disability earnings and where the statutory minimum wage equals 50% of pre-disability earnings. For this setting, Figure 3 shows post-application earnings, current benefits and total income as a function of post-application earnings. As the benefit system thresholds are based on percentages of pre-disability earnings, all amounts are shown as percentages of pre-disability earnings. The vertical dashed lines demarcate the benefit thresholds: relative earnings that are necessary to receive the wage subsidy (25%), relative earnings at which the benefits are reduced (50%) and the maximum earnings that are allowed together with DI receipt (65%).

As long as earnings are below 50% of the remaining earning capacity, benefits are linked to the minimum wage and any increases in earnings are not tapered by reductions in disability benefits. Once post-application earnings exceed 50% of the remaining earnings capacity, benefits are linked to pre-disability earnings and amount to 70% of the difference between pre-disability earnings and remaining earning capacity. Since

Figure 3: Income as a function of post-application earnings as a percentage of pre-application earnings, given an assessed disability degree of 50%.



disability benefits are linked to the pre-disability earnings instead of the minimum wage, there is a jump in the amount of benefits at 50% of the remaining earning capacity. This jump represents a strong financial incentive to exploit at least 50% of the remaining earning capacity. Also, note that the size of the jump is increasing in pre-application earnings. The amount of DI benefits remains constant until the full remaining earning capacity is used.[8]

For earnings exceeding 50% of the pre-disability earnings, disability benefits are taxed away at a 70% rate. Once the 65%-threshold is reached, disability benefits are terminated. This creates a drop in total income at 65% of pre-disability earnings, resulting in strong disincentives to increase earnings beyond 65% of pre-application earnings. The magnitude of this drop does not depend of the degree of disability. Total income drops by 24.5% (70% of 35%) of pre-application income from 89.5% of pre-disability earnings to 65% of pre-disability earnings.

For individuals who have not been awarded disability benefits, total income is equal to the dotted 45-degrees line in Figure 3, implying there is no implicit taxation. By comparing the total income of individuals with and without disability benefits, three

[8]In the example of Figure 3, the remaining earning capacity is 50% of the pre-application earnings. 50% of the remaining earning capacity thus equals 25% of the pre-application earnings. At current income between 25% and 50%, benefits thus amount to 35% of pre-application earnings (70% minus 70% of 50%=35%).

main effects of disability benefits emerge. First, there is an income effect as long as post-application earnings are below 65% of pre-disability earnings. This income effect likely implies that recipients work less than they would have worked in the absence of benefits. Second, there is an incentive effect for disability benefit recipients to earn at least 50% of their remaining earnings capacity. Lastly, there is a strong disincentive to earn more than 65% of pre-disability earnings.

In light the above-mentioned incentives, the response to recovery (defined as an increase in the potential hourly wage) of those with disability benefits may be smaller than the response to recovery of those without disability benefits. First, the income effect described above implies that at higher hourly wages, those with disability benefits will work less. Second, the incentive to earn at least 50% of the remaining earning capacity relaxes slightly, as fewer hours are needed to earn this amount. Lastly, return to full-time work is highly beneficial for those without benefits but only marginally for those with benefits. Specifically, the transition for 65% to 100% of full time would yield an income gain of 10.5 percentage-points of the pre-disability wage earnings only. So the overall expectation is that the joint incentives reduce the response to recovery for recipients, relative to non-recipients.

3 Data sources and sample selection

We link three data sources to analyze recovery of DI recipients: disability insurance application data (provided by UWV), monthly income data (Statistics Netherlands) and mental health treatment data (Statistics Netherlands). All data are administrative and cover all Dutch citizens. The following subsections discuss the various data sources and the sample selection.

3.1 Disability insurance application data

The disability insurance application data comprises all applications between January 2004 and June 2017, which gives a total of 670,171 applications. The data contain all information for both awarded and rejected applicants that is needed determine their earnings capacities and their degrees of disability. It includes the pre-application hourly wage and number of hours worked, as well as the post-application potential hourly wage and number of working hours.[9] Additionally, it includes the timing and the outcome of the award decision by UWV. The decision could be made before or after the formal start of benefit entitlement, so benefits can be awarded retroactively or proactively. The decision date is therefore a proxy for the actual start of benefit payments.[10]

Several health-related variables are included in the DI application data. The first group of variables concerns the medical diagnoses of the applicants. The diagnoses are either classified by so-called “CAS-codes” or by a categorization created by UWV. The CAS-codes are used by health and safety doctors responsible for the reintegration process of long-term absent workers of firms and are more detailed than the UWV codes.[11] The CAS-codes are available for 80% of all applications, whereas the UWV codes are available for 98% of all applications. The analysis will, where possible, combine both types of diagnoses information.

The application data also contains information on the type and number of functional limitations of applicants, as assessed by a medical assessor of UWV. These limitations range from physical limitations, such as neck-movement and use of hands, to limitations such as cognitive functions and work stress. There are a total of 17 limitation groups, and the severity of every limitation can range from zero, implying no limitation, to 7, implying a severe limitation. The functional limitations are used when determining the potential hourly wage an applicant could earn. The last health-related variable concerns the probability of improvement, as assessed by UWV (“reasonable to good”, “small” or “non-existent”).

3.2 Income data

Income data is acquired from two sources of Statistics Netherlands. The first one contains information on all employment contracts in the Netherlands between 2006 and

[9] Some individuals are deemed to be (fully) incapable for work by a medical examiner, based solely on medical grounds. The assessment of the remaining earnings capacity is not conducted for these individuals and their degree of disability is not stated in the application information. Unfortunately, we lack information that would enable us to distinguish these applicants from those that terminate their application before the actual assessment.

[10] As will be shown below, the proxy is accurate as benefit receipt changes strongly in the month of the decision.

[11] CAS-codes consist of a diagnose group letter, e.g. “P” for psychological diagnoses, and a three-digit number indicating the specific diagnosis (32 in total). Applicants can have at most thee CAS-codes and three UWV diagnoses group codes.

2018, including an individuals' monthly earnings, hours worked and monthly employment status. We combine multiple employment contracts registered at one and the same month, so as to obtain the total earnings and total number of hours worked. The employment indicator indicates whether an individual worked for at least one hour in a specific month. Hourly wages are obtained by dividing the monthly earnings by the monthly number of hours worked.[12] The second source of information concern monthly indicators for various forms of income for all individuals in the Netherlands from 1999 till 2016. The included categories are: employment, self-employment, unemployment benefits, social assistance, old age pensions, disability or illness benefits and other social services.[13] The disability or illness benefits indicator does not discriminate between disability and illness.[14]

We enrich the income data with administrative records from Statistics Netherlands on the year of birth, gender, nationality and level of education. Using household identifiers we are also able to link individuals to their partners, which is important to determine welfare eligibility.

3.3 Healthcare data

Throughout the analysis, two separate data sources on medical treatments will be used. The first source concerns data on mental health treatments. These data are derived from so called "Diagnosis Treatment Trajectories" (DBC's) that are used as payment units for complete medical treatments. DBC's comprise all treatments that are deemed necessary to alleviate or solve health problems. We observe all DBC's regarding mental health between 2011 and 2016. Mental health problems in the sample can be considered as severe, with average cost of treatment of approximately €5,500 and resembling roughly 140 hours of treatment. All DBC entries state the starting and ending date of treatment, even if this start or end is before or after the DBC period (if treatments lasted for more than one year). For individuals with multiple treatment trajectories, we

[12]The monthly earnings data does not discriminate between regular wage pay and bonus pay, causing some outliers in the hourly wages. Approximately 200 out of the 2,000,000 hourly wage entries (0.01%) exceed €1,000 per hour. These outliers are excluded from the analysis.

[13]See Figure A.1 in the Appendix for the evolution of income receipt of the sample

[14]Both income data sources contain information on employment status. The differences between the two sources is very small (at most 1% of the monthly employment indicators differ). Due to the fact that the first data source contains more recent years and has more detailed employment information, these data will be used.

Table 1: Sample selection criteria

Inclusion criteria	Remaining sample
All DI applications between 01-01-2004 and 01-07-2017	670,171
First applications only	605,757
Application after 01-01-2006	498,810
Application due to mental health problem	191,056
Temporarily and partially disabled (0%-80%)	116,403
Recovered before 01-01-2016	45,966
Recovered after application	25,080

use the earliest start date and latest end date. Unfortunately, information on whether an individual actually recovered because of the treatment is not available. The end of treatment should therefore be considered as a proxy for recovery.

The second data source concerns the yearly healthcare expenditures covered by basic health insurance for the years 2009 until 2017. Basic health insurance is compulsory in the Netherlands and covers the vast majority of all healthcare. The data shows the spending on various subcategories. We construct measures on mental healthcare expenditures and non-mental healthcare expenditures (see Appendix A.1). Using the healthcare expenditures in the pre-application waiting period as a baseline, we create proxies for recovery based on a substantial drop in the healthcare expenditures. While our results appear fairly robust to alternative proxies, our preferred measure uses the year before the application date (as all DI applicants have been disabled during this entire year) as a baseline measure and proxies the end of treatment year as the year in which healthcare expenditures drop below (and stay below) 20% of the baseline. Given that the expenditure data is only available on an annual basis, there is more measurement error in the proxy compared to the proxy based on the DBC data.

3.4 Sample selection

Merging the three data sources yields a sample of disability insurance applicants for whom the application information, mental health information, and employment and income history is observed. A selection was made to make the sample suitable for analysis. Table 1 illustrates the various sample selection steps that are taken.

The DI application data comprises 670,171 applications filed between January 2004 and July 2017. Re-examinations are usually stem from changes in labor supply or health status, so we restrict the sample to first applications only. Given that employment data is available since 2006, we only consider applications filed in 2006 or later. To ensure that the mental health problems are sufficiently severe to affect the employment status of individuals, only those who have applied for disability insurance due to some form of mental health diagnosis are selected. This then yields a sample of approximately 40% of all applications (191,056 individuals). As the analysis focuses on those who are expected to recover, we exclude applicants that are deemed fully and permanently disabled.

We next merge the application data to the mental healthcare data. As the mental healthcare data is only available until 2016, individuals are selected for whom the end of treatment occurred before the first of January 2016; this ensures that a new mental healthcare trajectory does not start shortly after the observed end of treatment. As a final selection step, we exclude individuals for whom the end of treatment occurs before the application date. The final sample then consists of 25,080 DI applicants for whom we have a proxy for mental health recovery. Descriptive statistics of this final sample are shown in the next section.

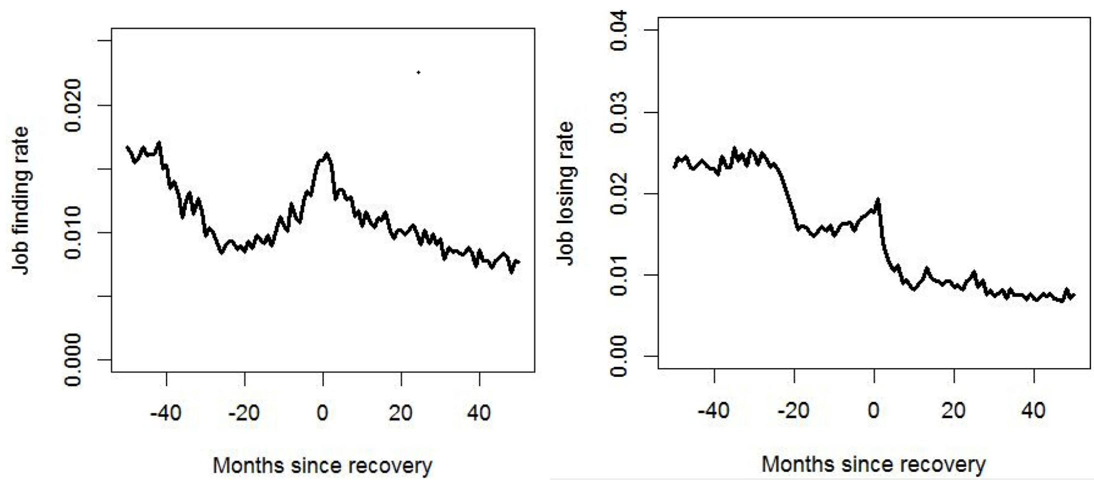
To gauge whether the end of mental health treatment indeed proxies recovery in terms of labor supply, Figure 4 shows the hazard rates for finding and losing a job around the moment of recovery. The job finding rate increases in the months before and after the end of the mental health treatment and the job losing rate drops after recovery. This suggests that the end of a mental health treatment indeed signals an improvement in the employment status for the selected sample. The hazard rates also indicate the presence of measurement errors and/or anticipation seems, with increases in the job finding rate already before formal recovery. As we will argue later on, this calls for estimation methods that control for reporting errors and anticipation effects.

4 Empirical Analysis

4.1 Model specification

For causal inference on the disincentive effects of disability benefits on the employment response to mental health improvement, we use Difference-in-Difference (DiD) specifications. In this framework, a control group is used to estimate the counterfactual for the treatment group (Lechner et al., 2011). Under the assumption that the

Figure 4: Job finding rate (left) and job losing rate (right) around recovery



control group and the treatment group follow parallel trends in outcomes variables in the absence of treatment, any divergence between the groups can be attributed to the causal impact of the treatment. In the context of our analysis, treatment is defined as recovery while receiving disability benefits, whereas the control case is recovery while not receiving benefits. The resulting model specification is as follows:

$$E_{it} = \alpha_1 + \alpha_2 \mathbb{1}_{DI_i} + \beta_1 \mathbb{1}_{t>R_i} + \beta_2 \mathbb{1}_{DI_i} \mathbb{1}_{t>R_{it}} + k(t) + \theta X_{it} + \varepsilon_{it}, \tag{2}$$

in which $\mathbb{1}_{DI_i}$ is an indicator function for receiving disability benefits and R_i is the time at which individual i finishes their mental health treatment. All DiD specifications we will use allow for time-constant pre-recovery level differences between the groups with and without benefits (α_1 and α_2) and flexible time trends $k(t)$ for which we will use different specifications. The effect of recovery on the control group is given by β_1. The parameter of interest is β_2, which is the difference between the effects of recovery for the groups with and without disability benefits.

In our DiD model, parameter β_2 describes the difference in recovery effects between those with and without DI benefits. At this point, it is important to stress that this difference consist of two components. First, the effect may be due to the benefit disincentive. Second, the effect may come from response differences inherent with the initial health differences. Even though our research strategy aims to construct treatment and control groups with similar health levels, applicants with higher disability degrees may have more severe mental health conditions and, consequently, less strong recovery effects. Consistent estimation of the disincentive effect therefore requires that these effects are small or that we control for them. In this respect, we characterize β_2 as an "unadjusted" DiD estimate.

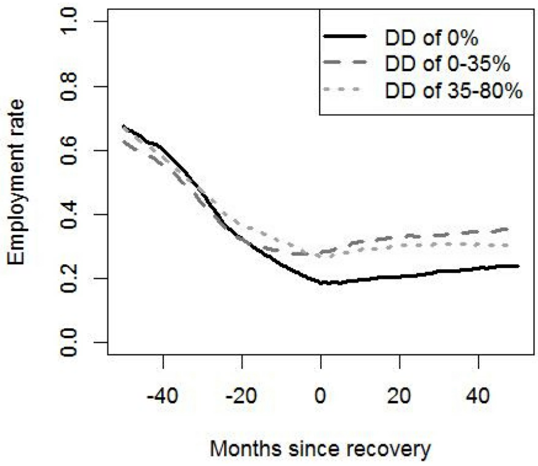
To adjust for the effect of differences in health on the response to recovery, we conduct DiD regressions on groups with the same DI award status (treatment and control group were both awarded benefits or both were rejected benefits) but different disability degrees. Consequently, any differences in recovery effects that follow from these comparisons cannot be attributed to benefit incentive effects and stem only from recovery response differences relating to differences in health. These resulting estimates can be subtracted from the unadjusted DiD estimates, so as to obtain the "adjusted" DiD estimate that measures the discouraging impact of benefits.

We will estimate Equation (2) on employment, hours worked and hourly wages as relevant outcome measures. In light of the longitudinal nature of the data and persistence in the outcome variables, we cluster standard errors by individuals (Hausman & Rapson, 2018).[15] Based on the graphical evidence shown below, a time window of 12 months before and 12 months after recovery is used in the baseline model. Both the pre-recovery and post-recovery time trends are assumed to be linear, while controlling for age, gender, nationality, education and include month and year dummies. As a robustness check, the linear time trend is replaced by monthly dummies (i.e. one month before recovery, two months before recovery, etc.).

Inherent with the DiD-framework, our estimation strategy relies on parallel trends of control and treatment groups. Specifically, parallel trends should apply for the specific group of individuals with and without DI benefits that end a mental health treatment at some point in time. It should be stressed that the parallel trend assumption does not impose a non-anticipation assumption here. As long as recovery is anticipated in a similar way by those with and without benefits, the parallel-trends assumption needed for identification of the disincentive effect is not violated. Next to this, another specific feature of our estimation approach concerns the use of the end

[15] As a robustness test, two additional methods will be used to account for the potential presence of serial dependence: (i) the model will also be estimated on mean levels and (ii) the analyses will be conducted non-parametrically by not including any control variables in the regressions and thereby comparing the differences in unconditional means. The latter should circumvent any time-series characteristic issues such as heterogeneity (Lechner et al., 2011).

Figure 5: Employment rates relative to recovery for DI applicant categories with disability degrees (DD) equal to 0%, 0-35% and 35-80%.



of mental treatments as a proxy for recovery. It is not likely that all individuals in the treatment and control groups have (fully) recovered after the end of treatment. Accordingly, the estimates of recovery effects can be interpreted as lower bounds or Intention-To-Treat effects.[16] For the consistent estimation of differences in recovery response effects (β_2) we therefore need the assumption to hold that the proportion of individuals that recovers is similar in the treatment and control group. So in addition to the parallel trends assumption, our estimation approach relies on the "equal-occurrence-of-recovery" assumption.[17] In what follows, we will analyze the validity of these two assumptions.

4.2 Parallel trends

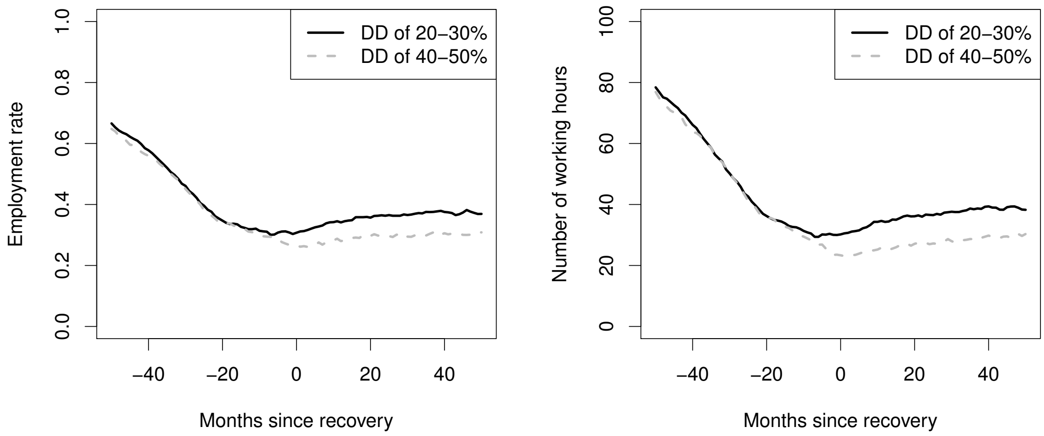
A natural starting point for selecting the control and treatment groups would be to consider the disability degree classes. In line with this, we initially divide our sample in three groups: the non-disabled (assessed disability degree of 0%, 9,502 applicants), the partially disabled without disability benefits (assessed disability degrees of 0%-35%; 9,489 applicants) and the partially disabled with disability benefits (assessed disability degrees of 35%-80%; 6,078 applicants).[18] To gauge the presence of parallel trends with

[16] If a share of ρ recovers in both the treatment and control group, the actual disincentive effects from disability benefits on the labor response to mental health recovery equals $\frac{\beta_2}{\rho}$.

[17] The following subsections focus on these assumptions for the unadjusted DiD. Results for the adjustment DiD are comparable and available on request.

[18] The non-disabled are treated as a separate group since their application outcome can be considered a corner solution: their remaining earnings capacity is assessed as at least as much as their pre-disability

Figure 6: Labour supply relative to mental health recovery for DI applicants with disability degrees (DD) of 20-30% vs. 40-50%.



(a) Employment (b) Hours worked

these three groups, Figure 5 shows the employment rates of the three groups, measured relative to the moment of recovery. All three groups show a deterioration of the employment rate in the years leading up to recovery and a stabilization after recovery, but the pre-recovery trends differ substantially. This observation is not surprising, since individuals with a loss of earnings capacity close to 80% are very different from individuals with a loss of earnings capacity close to 0%.[19]

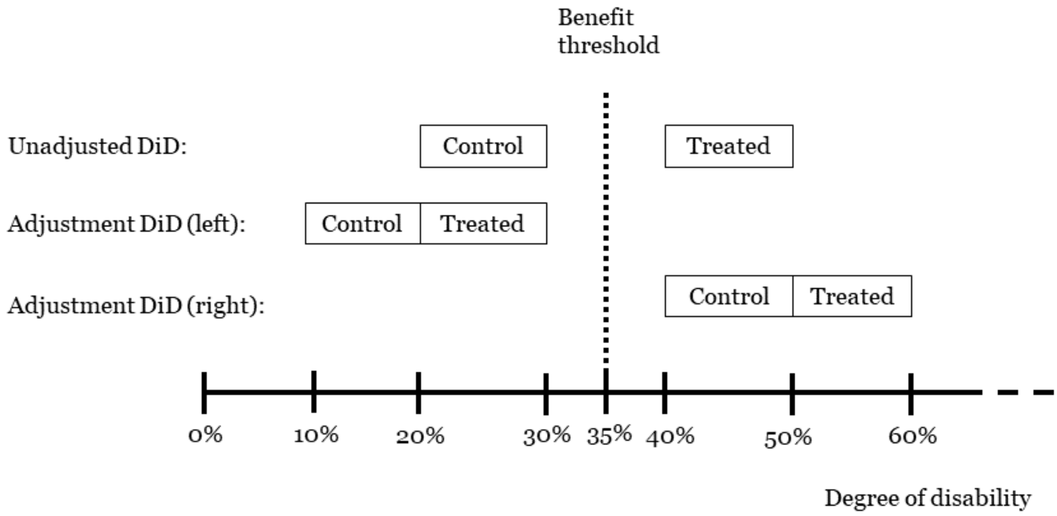
In light of these considerations, we select applicants which are closer to the 35% disability benefit cut-off to obtain groups that are more similar. In doing so, care should then be taken of potential manipulation effects around these cut-offs that are suggested by McCrary tests on the densities of observations (see Figure A.2 in the Appendix). Since assessors have some leeway to affect disability degrees, we therefore use a "donut"-design around the threshold. Specifically, in our baseline model we compare applicants with an assessed loss in earnings capacity of 20-30% (the control group) to applicants with an assessed loss in earnings capacity of 40-50% (the treatment group).[20] Panels (a) and (b) of Figure 6 show pre-recovery trends and levels are very similar for the control and treatment group, both for employment and hours worked

earnings.

[19]Figure A.2 in the Appendix shows the main characteristics of the three groups.

[20]As a robustness check we also consider larger bandwidths and we perform the analysis without the donut around the threshold, using the 25-35% group as control group and the 35-45% as treatment group. Results are very similar, as we show in Table 4.

Figure 7: Derivation of unadjusted and adjustment DiD disincentive effects

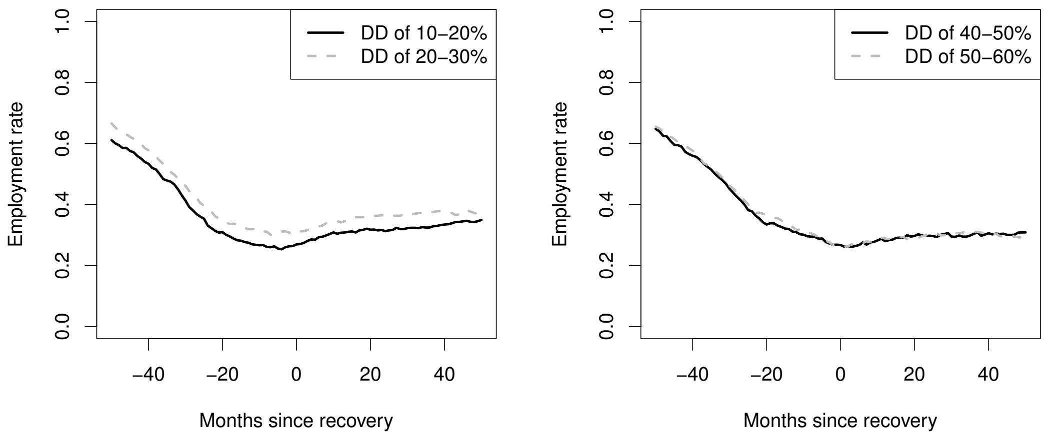


as outcome variables. Figure 9 shows that the evidence of parallel trends appears less strong for hourly wages. While the wage patterns follow the same general trend, there is more noise. Inherent with the fact that larger wages are associated with lower disability degrees, the average pre-recovery wage level is higher for those with benefits.

We explained earlier that the calculation of adjusted DiD estimates requires the outcomes of DiD analyses of groups with similar DI benefit statuses. With 20-30% and 40-50% disability degree-groups to obtain the unadjusted DiD estimates, we consider neighboring groups. Specifically, we use the 10-20% group as a control group for the 20-30% group and the 50-60% group as control group for the 40-50% group. Figure 7 illustrates how these outcomes can be used to obtain the adjusted DID estimates. In addition, Panels (a) and (b) of Figure 8 display the parallel trends in employment for the comparisons of neighboring groups with similar award statuses. The trends in employment – and number of working hours (not shown here) – for both adjustments appear well in tandem before and after the moment of recovery.

We further assess the parallel trends assumption by comparing characteristics of the control and treatment group. Table 2 shows descriptive statistics for the 20-30% and 40-50% groups and tests for equality of means. Both groups have similar gender and nationality statistics, but the control group without benefits is on average slightly younger and lower educated. The length of mental health treatment is approximately

Figure 8: Employment relative to recovery from mental health for DiD adjustment groups



(a) 10-20% vs. 20-30% groups

(b) 40-50% vs. 50-60% groups

two and a half months shorter for those without benefits. As expected, there are significant differences in the DI application variables. The number of functional limitations, the degree of disability, the pre-application hourly wage and the pre-application number of working hours are all lower for those without benefits. The assessed probability of health improvement at the moment of application is very similar for both groups. Lastly, both groups have similar healthcare expenditures in the year of their DI application. The overall picture that emerges is that there are statistically significant differences between the two groups, but these are limited in substance.[21]

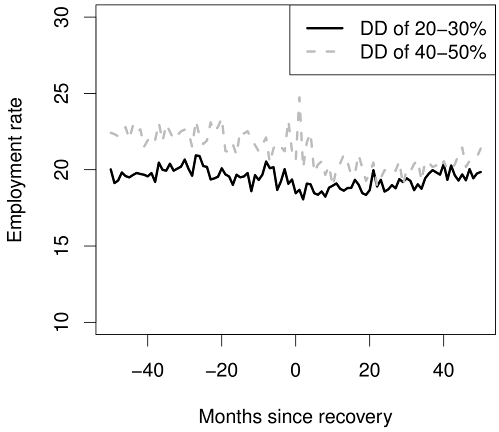
4.3 Equal occurrence of recovery

We stated earlier that recovery effects in our model can be interpreted as Intention-to-Treat effects. That is, we interpret the termination of a mental health treatment program as a signal of mental health improvement, but is does not necessarily imply full recovery. We therefore need the assumption to hold that this proxy is equally strong in the control and treatment group. We consider this issue in more detail.

First, one might worry that the outcome of the award decision affects the subsequent duration of the mental health treatment. If so, the end of a mental health

[21]This is confirmed in one of the robustness tests to our model, where we show that the inclusion of individual control variables has a negligible impact on our findings.

Figure 9: Average hourly wage relative to recovery from mental health for DI applicants with disability degrees (DD) of 20-30% vs. 40-50%.



treatment may have implications that differ between the group with and without benefits. To provide insight in the importance of such effects, Panel (a) of Figure 10 shows the Kaplan-Meier estimates of the treatment duration since the time of application. It is reassuring to see that the survival functions are very similar, with a log rank test indicating insignificant group differentials (P-value = 0.70).

A second concern is that the actual, underlying rate of recovery differs between the control and treatment group. Since actual recovery is unobserved, we need to make comparisons on alternative recovery measures that provide indirect inference on this matter. Specifically, we observe that the probability of improvement that is assessed by UWV (Table 2) is very similar for the treatment and control groups. In a similar vein, we can compare the duration of the mental health treatment programs. For example, if reverse causality is stronger in one of the groups – meaning more people recover *because of* working) – the duration of treatments should be shorter on average for that group. To test for this, Panel (b) of Figure 10 show the fractions of individuals still receiving mental health treatment as a function of treatment duration for both groups. A log rank test shows no significant difference between the survival functions. We therefore conclude that recovery rates due to mental health treatment and due to reverse causality are comparable in size for the control and treatment group.[22]

[22]Note that similar results are obtained for the disability-degree-groups that we use to adjust our DiD-estimates. Results are available on request.

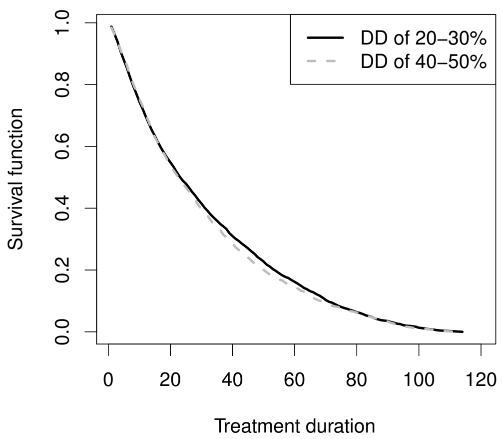
Table 2: Descriptive statistics of DiD treatment and control groups

	Degree of disability:		
	20-30%	40-50%	P-val[a]
Demographics:			
Age	47.565	49.013	0.000
Female	0.538	0.549	0.474
Dutch native	0.655	0.692	0.010
Education:			
Unknown	0.087	0.138	0.000
Low	0.269	0.239	0.022
Middle	0.437	0.367	0.00
High	0.207	0.257	0.000
Employment:			
Pre-application hourly wage	15.703	17.718	0.000
Pre-application hours	33.087	34.025	0.003
Mental health:			
Treatment duration[b]	32.935	35.317	0.000
DI application:			
FML	9.685	11.613	0.000
Disability percentage	25.204	44.610	0.000
Earnings capacity: hourly wage	12.211	12.096	0.023
Earnings capacity: hours	32.420	27.110	0.000
Assessed probability of health improvement:			
NA	0.278	0.258	0.130
Reasonable to good	0.641	0.679	0.007
Small	0.078	0.059	0.011
Non-existent	0.003	0.004	0.617
Medical expenditures:			
Mental healthcare expenditures[c]	3,037.046	3,542.620	0.136
Physical healthcare expenditures[c]	1,427.917	1,611.686	0.095
Observations	3,346	1,656	

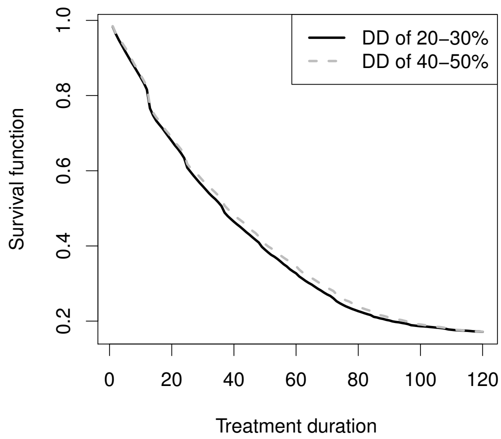
[a]P-value of two-sample t-test for equality of means; [b]Duration of the mental health treatment in months; [c]Expenditures in euros in the year of DI application.

For the estimation results that follow, we will also investigate the equal-recovery assumption by estimating DiD models on sub-samples of individuals with specific mental health problems: those with mood disorders, anxiety disorders or personality disorders. We hypothesize that the estimates for anxiety disorders and personality disorders are less susceptible to reverse causality, as there is no empirical evidence for such a relationship in the literature (Paul & Moser, 2009).

Figure 10: Survival functions of mental health treatment duration for DI applicants with disability degrees (DD) of 20-30% vs. 40-50%.



(a) Time of recovery, relative to application



(b) Duration of mental health treatment

4.4 Main estimation results

Table 3 shows the unadjusted estimates, the adjustment estimates and the resulting adjusted estimates of the baseline DiD model in Equation (2). Recall from Figure 7 that the adjustment follows from the DiD estimates on groups with similar award outcomes. At the left-hand-side, this entails a comparison between the groups with disability degrees of 20-30% and 10-20%. At the right-hand-side, we compare the 40-50% group to those with 50-60%. These comparisons yield DiD results that we use to obtain the "adjusted" estimates that control for potential differences in health conditions between the disability degree groups. It should be noticed that the average difference in degree of disability for the unadjusted DiD estimate is 20 percentage-points, whereas the average difference is 10 percentage-points for the adjustments. We therefore deduct both adjustments to obtain the adjusted estimates.

In line with expectations, the adjusted DiD estimates for the effect on employment and hours worked are somewhat smaller than the unadjusted DiD estimates. The estimated unadjusted treatment effect on employment is 3.5 percentage-points (Panel A, row 1), whereas the adjusted estimate shows that receiving DI benefits reduces the effect of recovery on employment by 1.8 percentage-points, compared to those who do not receive disability benefits. The disincentive effect on the number of hours worked

Table 3: DiD estimates for employment, working hours and hourly wage

Panel A: Employment

	Unadjusted estimate[a]	Left-hand-side adjustment[b]	Right-hand-side adjustment[c]	Adjusted estimate[d]
DiD estimate	-0.035**	-0.010**	-0.007*	-0.018**
	(0.002)	(0.002)	(0.002)	(0.003)[e]
Pre-recovery difference	-0.031**	0.038**	0.010*	
	(0.002)	(0.002)	(0.003)	
Recovery	0.007**	0.006**	-0.004	
	(0.001)	(0.001)	(0.001)	
Pre-recovery trend	-0.001**	0.000	-0.004**	
	(0.000)	(0.000)	(0.000)	
Post-recovery trend	0.003**	0.005**	0.002**	
	(0.000)	(0.000)	(0.000)	

Panel B: Monthly number of working hours

	Unadjusted estimate[a]	Left-hand-side adjustment[b]	Right-hand-side adjustment[c]	Adjusted estimate[d]
DiD estimate	-4.984**	-0.626*	-1.591**	-2.767**
	(0.207)	(0.193)	(0.248)	(0.376)[e]
Pre-recovery difference	-4.288**	4.168**	1.679**	
	(0.274)	(0.236)	(0.325)	
Recovery	1.655**	0.682**	-0.020	
	(0.111)	(0.143)	(0.155)	
Pre-recovery trend	-0.340**	-0.114**	-0.689**	
	(0.015)	(0.013)	(.018)	
Post-recovery trend	0.328**	-0.469**	0.229**	
	(0.014)	(0.014)	(0.016)	

Panel C: Hourly wage

	Unadjusted estimate[a]	Left-hand-side adjustment[b]	Right-hand-side adjustment[c]	Adjusted estimate[d]
DiD estimate	0.225	-0.450*	0.023	0.652
	(0.252)	(0.227)	(0.270)	(0.434)[e]
Pre-recovery difference	1.283**	2.184**	0.316	
	(0.229)	(0.232)	(0.254)	
Recovery	-0.185	-0.044	1.150*	
	(0.230)	(0.227)	(0.298)	
Pre-recovery trend	0.023	0.015	0.057*	
	(0.017)	(0.014)	(0.025)	
Post-recovery trend	-0.045*	-0.020	-0.144**	
	(0.016)	(0.013)	(0.025)	
Sample size	5,002	5,943	3,149	

[a]Comparison between 20-30% and 40-50% disability degree groups; [b]Comparison between 10-20% and 20-30% disability degree groups; [c]Comparison between 40-50% and 50-60% disability degree groups; [d]The adjusted estimate is computed as the unadjusted estimate minus the left-hand-side and the right-hand-side adjustment. [e] Assuming independence, the standard error of the adjusted estimate is computed as $\sqrt{\mathrm{sd}_a^2 + \mathrm{sd}_b^2 + \mathrm{sd}_c^2}$

Standard errors are shown in parentheses. *significant at a 5% significance level, **significant after applying a $\frac{1}{60}$ Bonferroni correction factor.

entails a reduction of 2.8 hours (Panel B, row 1). The employment and hours effects are statistically significant, also after applying Bonferroni correction for multiple hypothesis testing.[23] Finally, the disincentive effect is positive but statistically insignificant for hourly wages (Panel C, row 1).

To assess the magnitude of disincentive effects, we next perform back-of-the-envelope calculations that compare the actual recovery effects of individuals with and without benefits. For this, we use the adjusted DiD estimates. We assume that the employment rates would stabilize in the absence of recovery. Based on the estimates of the post-recovery trends and the initial recovery effect, we first calculate recovery effects that accrue over 12 months for the 20-30 group (individuals without benefits). The instantaneous recovery effect on employment for this control group equals 0.7 percentage-points, whereas the post-recovery time trend accumulates to 3.6 percentage-points after 12 months. This yields a 12-months-recovery effect on employment equal to 4.3 percentage-points. The disincentive effect of benefits receipt (1.8 percentage-points) thus largely offsets this increase in employment (approximately 40%). Using similar calculations for the 12-months recovery effect on hours worked, we derive values of 5.6 and 2.8 for those without and with benefits. Again, this indicates that approximately half of the recovery effect is offset by the disincentives of partial DI benefits.

4.5 Robustness analyses

Table 4 presents the results of various robustness analyses to our baseline model. The analyses consider the effect of changes in the time window around recovery, the exclusion of observations around recovery and changes in the samples of treatment and control groups. Table A.3 in the Appendix complements these findings with robustness tests that yield results that are very similar to those for the baseline model.[24] For expositional reasons, both tables report results for the unadjusted DiD estimates of the disincentive effect only. Albeit that the size of the adjusted coefficients is mostly smaller for the outcome variables, we also find significant coefficients in all cases where

[23] We use a $\frac{1}{60}$ Bonferroni correction factor as we test a total of 60 DiD estimates. Given that the estimates of robustness specifications are not independent, the correction factor that is used can be considered as conservative.

[24] These robustness tests include models without covariates, the use of monthly dummies, mean-level estimation and non-parametric estimation.

Table 4: Robustness analyses for the unadjusted[a] DiD effects for employment, working hours and hourly wage as outcome measures

	Specification				Outcome measures		
	Window[b]	Donut	N_c^c	N_t^d	Employment	Hours	Wage[e]
Baseline model	12	0	3,346	1,656	-0.035**	-4.984 **	0.225
					(0.002)	(0.207)	(0.252)
6-month time window	6	0	3,346	1,656	-0.022**	-3.280**	1.227*
					(0.002)	(0.238)	(0.466)
48-month time window	48	0	3,346	1,656	-0.053**	-7.786**	-1.116**
					(0.001)	(0.141)	(0.0.063)
Donut	48	12	3,346	1,656	-0.059**	-8.760**	-1.505**
					(0.001)	(0.193)	(0.074)
15-30% vs. 40-55%	12	0	4,726	2,480	-0.041**	-5.412**	0.032
					(0.002)	(0.173)	(0.206)
25-35% vs. 35-45%	12	0	3,270	1,788	-0.015**	-2.650**	0.660*
					(0.002)	(0.207)	(0.312)

[a]Comparison between 20-30% and 40-50% disability degree groups; [b]Incorporated number of months before and after recovery; [c]Number of individuals in the control group; [d]Number of individuals in the treatment group; [e]Hourly wage; Standard errors shown in parentheses; *significant at a 5% significance level; **significant after applying a $\frac{1}{60}$ Bonferroni correction factor.

unadjusted coefficients are statistically significant as well.[25]

Generally, our DiD estimates for employment and number of working hours in Table 4 are robust to the various specifications used. Shortening the time window around the moment of recovery to six months decreases the DiD estimates to some extent (row 2). Presumably, this reflects the fact that a relatively large part of the pre-recovery divergence in outcomes is attributed to the pre-recovery difference in means, which leaves less room for post-recovery treatment effects. Using similar arguments, expanding the time window to 48 months increases the effect on employment and hours to 5.3% points and 7.8 hours, respectively (row 3). We next consider a symmetric donut specification that excludes observations within 12 months around the moment of recovery (row 4). Since this specification attributes any divergences within these 12 months to recovery, the resulting DiD estimates provide upper-bounds. The estimated effect on employment is 5.9% points and the effect on number of working hours is 8.8 hours; these effects correspond to changes of 20% and 25%, respectively. Given the large effect on working hours, this mostly reflects both an intensive and extensive margin effect. Finally, our estimates are robust to changes in the samples of treatment and control groups. For this we consider groups further away from the 35% disability threshold (row 5) and (smaller) groups closer to the disability threshold (row 6). The

[25]Robustness results for the adjustment estimates are available on request.

Table 5: Heterogeneity by mental disorder type: unadjusted[a] DiD estimates for employment, hours and hourly wage

	Specification				Outcome measures		
	Window[b]	Donut	N_c^c	N_t^d	Employment	Hours	Wage[e]
Baseline model	12	0	3,346	1,656	-0.035**	-4.984**	0.225
					(0.002)	(0.207)	(0.252)
Anxiety disorders	12	0	433	174	-0.031**	-5.038**	-0.753
					(0.006)	(0.666)	(0.401)
Personality disorders	12	0	556	279	-0.023**	-4.035	0.463
					(0.005)	(0.561)	(0.363)
Mood disorders	12	0	1,304	690	-0.040**	- 5.376**	0.135
					(0.003)	(0.312)	(0.523)

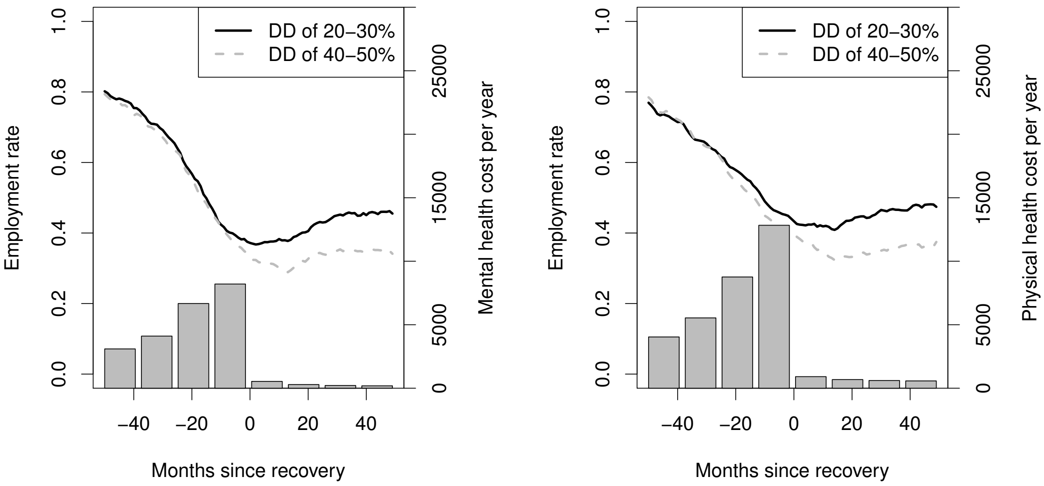
[a]Comparison between 20-30% and 40-50% disability degree groups; [b]Incorporated number of months before and after recovery; [c]Number of individuals in the control group; [d]Number of individuals in the treatment group; [e]Hourly wage. Standard errors are shown in parentheses; *significant at a 5% significance level; **significant after applying a $\frac{1}{60}$ Bonferroni correction factor.

results for hourly wages are not robust to the various specifications, both in terms of sign and significance.

To shed light on heterogeneous disincentive effects, Table 5 shows the results of the baseline models estimated on sub-samples that are stratified by types of mental health diagnoses. The estimates for the largest group, those diagnosed with mood disorders, are larger than the baseline estimates. As mentioned earlier, reverse causality could be relevant for this sub-sample. For the other two sub-samples reverse causality is less likely. The estimated effects on employment and number of hours worked for those with anxiety disorders and personality disorders are large and significant as well. Given the large effects for the sub-samples with anxiety and personality disorders, it is unlikely that our findings are (mainly) driven by reverse causality.

Summing up, our evidence points at significant disincentive effects of having disability benefits on the employment response to mental health recovery. The estimated effects range between 2.2 to 5.9 percentage-points in terms of employment and between 2 and 9 working hours per month. Part of these effects already seem to accumulate before the end of treatment, implying that individuals either anticipate their recovery or that recovery happens before the end of treatment. Since the relative effects on the average number of working hours exceed those on the employment rate, there is both an intensive and extensive margin effect on hours worked. Given the robustness of the results, it seems unlikely that the results are due to reverse causality.

Figure 11: Employment rates relative to recovery based on healthcare expenditures



(a) Mental health

(b) Physical health

4.6 Analysis with healthcare expenditure data

Next to mental health treatments, one may argue that changes in the yearly medical health expenditures and physical health expenditures may also indicate recovery. While yearly data provide us with estimates that are less precise than with monthly data, the advantage is that they allow us to study disincentive effects that extend to non-mental health. We therefore conduct similar analyses of recovery effects, but now with drops in physical and mental healthcare expenditures as proxies for recovery. Specifically, we define the year of recovery as the first year in which expenditures drop below and stay below 20% of the healthcare expenditures in the year before the DI application.[26] We distinguish between mental healthcare expenditures and non-mental healthcare expenditures (see Appendix A.1 for details), yielding two respective proxies for recovery. We select applicants for whom we observe physical or mental recovery and that have disability degrees in the relevant treatment and control groups that have been explained earlier. The resulting sample contains 7,418 individuals with mental health-recovery and 9,747 individuals with non-mental health-recovery.

For both samples and corresponding types of recovery, Figure 11 shows lines for the evolution of employment rates relative to the year of recovery (see the left Y-axis) as well as bars indicating the average mental or physical healthcare expenditures for each year (right Y-axis). Prior to recovery, expenditures increase to approximately

[26]Results for other thresholds are similar, see Table A.4 in the Appendix.

Table 6: Unadjusted[a] DiD estimates based on annual mental and physical healthcare expenditure for employment, hours worked and hourly wage

	Specification:						
	Window[b]	Donut	N_c^c	N_t^d	Employment	Hours	Wage[e]
(A): Mental health recovery							
Baseline specification	24	0	4,735	2,683	-0.065**	-9.996**	-1.019**
					(0.001)	(0.163)	(0.087)
Donut specification	24	12	4,735	2,683	-0.084**	-13.955**	-1.245**
					(0.003)	(0.313)	(0.153)
(B): Physical health recovery							
Baseline specification	24	0	6,301	3,446	-0.047**	-7.617**	-0.342*
					(0.001)	(0.138)	(0.073)
Donut specification	24	12	6,301	3,446	-0.071**	-11.346**	-0.597*
					(0.002)	(0.267)	(0.126)

[a]Comparison between 20-30% and 40-50% disability degree groups; [b]Incorporated number of months before and after recovery; [c]Number of individuals in the control group; [d]Number of individuals in the treatment group; [e]Hourly wage. Standard errors are shown in parentheses; *significant at a 5% significance level; **significant after applying a $\frac{1}{60}$ Bonferroni correction factor.

€8,000 and €13,000 per year. After the decrease, the expenditures remain low. The trends in the employment rates are very similar for the control and treatment group. A divergence in employment rates starts approximately one year before recovering, corresponding with the fact that recovery most likely occurs in the year prior to the first low-cost year. Before this year, the parallel trends assumption appears to hold. The divergence around recovery is significant for both proxies and the difference between the two groups remains relatively constant after recovery.

Table 6 displays the DiD estimates for both proxies for recovery that use the annual medical expenditure data. Panel A shows the results for recovery based on a decrease in mental health expenditures and Panel B shows the results for recovery based on a decrease in physical health expenditures. Similar to our earlier baseline model, we incorporate 24 months before and 24 months after recovery. Knowing that recovery could either occur in the first low-cost year or in the year prior to the first low-cost year, we also present outcomes where we incorporate a donut of 12 months before and 12 months after the moment of recovery. The estimates for mental health are larger than those obtained in our baseline results; this possibly reflects the fact mental health treatments are only a subset of all mental healthcare expenditures. The estimates for physical health are somewhat smaller in size than those for mental health. This either suggests that a drop in healthcare expenditures is less indicative of increases in the capacity to work than a drop in mental health expenditures, or that disincentive effects are less important for those with physical health improvements.

5 Labor supply effects in a structural model

Our estimation results so far point at distinct employment effects of recovery for individuals with and without DI benefits, both for health improvements that are proxied by the end of mental health treatments and by substantial drops in medical consumption. When interpreting these results, recall that the employment effects after recovery only partly compensate the large employment drop that has occurred in earlier years. We stated earlier that the response effects can be characterized as ITT-effects, since the end of the treatment implies the partial or full recovery only for a part of the relevant group. A pertaining question therefore is how large employment effects of full recovery are and what the maximum discouraging impact of DI benefits are.

To gain insight in the full effects of recovery, we develop a structural labor supply model that incorporates both health shocks and the subsequent recovery from health. With the data on the assessed earnings capacity and possible hours restrictions of workers at the moment of DI application, the impact of health changes can be modelled as changes in budget constraints and changes in the maximum number of hours that can be worked. Following Low and Pistaferri (2015), we model this as changes in utility preference parameters that stem from health shocks. These health shocks may permanently or temporarily change preferences for leisure. In effect, labor supply changes may stem both from productivity losses and a higher disutility from working.

To estimate the structural model parameters, we distinguish three successive stages that are relevant for the individual worker: (1) before the health shock, (2) after the health shock and (3) after recovery. For the first stage, we estimate utility preference parameters for work and leisure for each individual in the sample – based on the observed hours decision (and the hourly wage) before the onset of a disability. In the second stage, the onset of the disability implies a loss in earnings capacity along two observed dimensions: the maximum number of working hours and the hourly wage. If the loss of earnings capacity exceeds 35% of the previous wage earnings, DI benefits are awarded. Together with this decrease in the earnings capacity, we incorporate a shock to the disutility in the model. In the third stage recovery implies that the earnings capacity is restored and utility parameters return to their pre-disability level. Eligibility to DI benefits is maintained for those awarded benefits.

Our model abstracts from tax effects and assumes hourly wages to be exogenous. In addition, we assume that individual receive social assistance if their income from earnings and DI is below the social minimum. In the Netherlands, most workers are entitled to unemployment benefits in the short term, and to social assistance in the long term when UI benefits are exhausted. Our focus is therefore on the long term effects that occur after the onset of disability. The following subsection describes the specification of the model and presents a graphical illustration. Subsequently, we discuss the estimation of the model parameters and a counterfactual simulation.

5.1 Model setup

Our model assumes utility maximization over the number of hours worked. We adopt a Cobb-Douglas utility function with utility weights normalized to one.[27] Since eligibility to social assistance depends on partner income, our focus is on income at the household level. The general utility maximization problem for individual i is as follows:

$$\max_{E_i} u(L_i, I_i) = L_i^{\lambda_i+\delta_g} I_i^{1-\lambda_i-\delta_g} \tag{3}$$

$$\text{s.t. } L_i = T - E_i \tag{4}$$

$$I_i(E_i) = E_i w_i + DI(E_i) + \tilde{I}_i \quad \text{if} \quad E_i w_i + DI(E_i) + \tilde{I}_i \geq SA_i \tag{5}$$

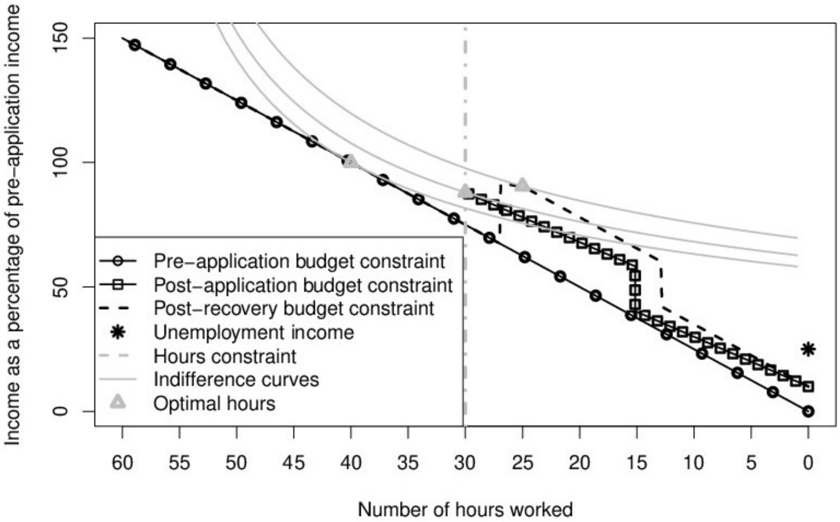
$$I_i(E_i) = (1-F)\; SA_i \quad \text{if} \quad E_i w_i + DI(E_i) + \tilde{I}_i < SA_i \tag{6}$$

with T the total amount of time an individual can divide between leisure L_i and employment E_i. We set $T = 60$, the highest observed pre-application number of working hours. I_i is total income, consisting of labor income $(E_i\, w_i)$, potential DI benefits $(DI(E_i))$ and partner income $(\tilde{I}_i)$. If income falls below the social assistance level, it is supplemented up to this level.[28] Income from social assistance is discounted by factor $1-F$, resembling the stigma attached to receiving social assistance. The size of F is identified from the size of hidden unemployment in the DI application inflow, i.e. the number of people for whom the pre-application utility level from working is below the pre-application utility level of receiving social assistance. Lastly, λ_i displays

[27]Since the utility function is estimated on a single employment decision, we can identify one preference parameter at most.

[28]The social assistance level is approximately €1,000 per month for singles and €1,500 for couples.

Figure 12: Budget constraints and utility indifference curves of a fictitious individual awarded partial disability benefits with a disability degree of 40%



the individual specific utility parameter and δ_g is a group-specific shock to this utility parameter. By construction, δ is equal to zero in the pre-application stage.

To illustrate the functioning of the model, Figure 12 shows the choice set for individuals with DI benefits.[29] The figure does not include any shocks to the utility function and demonstrates an individual working full-time (40 hours) before the DI application. The individual has an assessed degree of disability of 40% and will therefore be awarded disability benefits. The degree of disability is the result of an hours restriction of 75% and a drop in the hourly wage equal to 80%.

To start with, the pre-application budget line increases one-to-one with the wage earnings. The post-application budget constraint is indicated by the line with squares; it shows that the slope of the budget constraint decreases due to the 20% reduction in hourly wage. Furthermore, the disability benefits cause non-linearities at the various thresholds of the DI system. Most notably, there is an incentive to work at least half of one's remaining capacity (15 hours in this case) and the assessed disability degree enforces a maximum number of working hours such that the budget constraint ends at 30 hours in this case. Lastly, the post-recovery budget constraint is shown by the dark dashed line. As long as the individual receives disability benefits after recovery, the discontinuities of the post-application period persist. With a higher hourly wage, these discontinuities occur at a lower number of hours worked. In the example shown

[29]See Appendix A.3 for the same illustration for individuals without DI benefits.

above, the response to recovery therefore entails a further reduction in the number of working hours. This contrasts to the case of a similar individual without disability benefits, who would increase the number of working hours up to the pre-disability level (see Online Appendix A.6).

5.2 Estimation procedure

We briefly describe the model estimation and refer to Online Appendix B.2 for further details. Following the timing of the employment decisions an individual faces, the estimation occurs in three stages. In the pre-application step, we estimate the individual utility parameter δ_i using observations of the hourly wages and the number of working hours. Given that every individual in the sample worked prior to applying for DI benefits, there is a unique δ_i for each individual which results in the observed number of working hours. The fixed cost of social assistance income, F, is then determined by setting the percentage of individuals who prefer social assistance over their observed number of working hours to 5%.[30] The resulting fixed cost of receiving social assistance equals 0.463. This means one euro received from social assistance is valued the same as €0,54 earned through working.

In the second step, the individual has experienced a health shock and may or may not have been awarded benefits. The budget constraints are altered due to a reduction in potential hourly wage, a restriction on the maximum amount of working hours and the provision of DI benefits for those awarded DI benefits. Using these new budget constraints, the new optimal number of working hours are predicted and confronted with the observed post-application number of working hours. To optimize the fit to these data, the group-specific utility parameter shock (δ_j) is set such that the average predicted number of working hours of each group equals the average observed number of working hours. The resulting shock estimates are 0.24 for those without benefits and 0.26 for those with benefits.[31]

In the post-recovery step the response to recovery is simulated by restoring the earnings capacity to its pre-application level and setting the utility parameters to their original values as well (i.e., we set δ equal to zero). In this respect, note that

[30] As a robustness test, we used a a 10% level instead. This yielded very similar results for the benefit disincentive effects we obtain in the recovery-stage.

[31] Note that the similarity in shock effects indeed suggests groups in the DiD analysis are similar.

Table 7: Implied employment and working hours effect of DI benefit receipt for workers with disability degrees of 40-50%; counterfactual analysis based on structural model

	Post-application	Recovery of earnings capacity	Recovery of earnings capacity and utility parameters
Employment			
Without DI benefits	31.0%	54.1%	96.6%
With DI benefits	32.7%	50.8%	81.3%
DI benefit effect	1.7%	-3.3%	-15.3%
Weekly working hours			
Without DI benefits	5.2	10.0	33.3
With DI benefits	5.5	6.6	16.0
DI benefit effect	0.3	-3.4	-17.3

the differences in the predicted response between the 20-30% group and the 40-50% group can be interpreted as the analogue of the the unadjusted DiD estimate (as they also include the effect of differences in health). These results can be found in Online Appendix B.2.[32]

5.3 Counterfactual employment outcomes of awarded individuals

With a larger assessed loss in earnings capacity, individuals with DI benefits are by construction different from individuals without DI benefits. Similar to the reduced form estimates in the previous section, this means that derived differences in response rates to recovery may also reflect compositional differences. Given the structural setup of our model, however, we can offset such effects by constructing counterfactual employment outcomes for the sample of individuals with DI benefits, as if they had not been granted DI benefits instead.[33] This allows us to assess the discouraging impact that benefit receipt may have on work resumption after application and recovery from health. Table 7 shows the results that follow from this approach.[34]

For the post-application stage, we find employment rates that are comparable in

[32]The structural model also allows for computation of the adjusted DI benefit effect, as shown in the next subsection.

[33]A similar exercise was conducted for applicants without DI benefits, giving similar results.

[34]Given our structural parameters, we can also broaden our analysis to recovery effects with benefit conditions that differ from the Dutch context. In this respect, one may argue that incentive structures are different with different replacement rates or earnings caps and without minimum earnings requirements. In the Online Appendix B.2 we have conducted such analyses.

the 'true' scenario with DI benefits and in the fictitious scenario without DI benefits. The employment rate is even slightly lower in the absence of DI benefits. The predicted employment rates differ by 1.7 percentage-points and the predicted average number of working hours are almost identical.[35] This suggests that the discouraging impact of benefit receipt is small among disabled individuals.

Upon recovery of the earnings capacity, there is a widening in the difference in the employment rates of the two groups. Assuming that utility preference parameters stay constant but the earnings capacity is restored (as shown in the second column), DI benefits seem to discourage 3.3 percentage-point from work resumption, as compared to a maximum work resumption effect (without benefits) equal to about 23 percentage-points. When utility parameters are also restored to their pre-application values and there is "full" recovery (third column), the discouraging impact increases to 15.3 percentage-points. Roughly speaking, this is about a quarter of the impact of full recovery for those without benefits if about 65 percentage-points. For working hours, we also see a widening of disincentive effects after recovery of the earnings capacity and after full recovery. The relative size of these effects is larger, indicating that the structural model predicts decreases in working hours for those employed and receiving benefits. In this way, these individuals avoid the loss of DI benefits.

6 Conclusion

This paper studies whether labor supply responses to improvements in (mental) health are partly eliminated by the disincentives of disability benefits. In doing so, we deepen the understanding of low work resumption rates of DI benefit recipients, particularly in schemes intended for those deemed temporarily disabled. Applying a Difference-in-Differences (DiD) framework, we compare Dutch DI applicants below the disability degree eligibility threshold as a control group with those above the disability degree threshold. The control and treatment groups have parallel trends in employment leading up to recovery, as proxied by the end of mental health treatment. We also show that recovery rates are similar for both groups and we adjust our estimates for potential differences-in-health effects. Following this approach, we find that the

[35]Recall that the DI scheme inhibits an incentive to exploit at least half of the remaining earnings capacity. This may explain why the overall effect is negative.

disincentive effects of disability benefits amount to approximately half of the recovery effect for individuals without DI benefits.

As our reduced form results can be considered as Intention-To-Treat estimates that resemble the effect of partial recovery, we also construct and estimate a structural labor supply model to uncover full recovery effects. Key to this structural approach is that individuals with and without DI benefits both experience a full recovery of their earnings capacity, but have distinct budget constraints. We then find that disability benefits reduce the response to recovery by approximately 15 percentage-points in terms of employment, suggesting that DI benefits absorb at least a quarter of the response in case of full recovery.

Compared to earlier findings, the reduced form estimates of disincentive effects in our setting are limited to an employment reduction of about two percentage-points. In the literature, estimates of employment reductions typically range between 20 and 30 percentage-points (Bound, 1989; Chen & Van der Klaauw, 2008; Maestas et al., 2013; French & Song, 2014). We argue that there are two explanations for this difference. First, the end of mental health treatments represents only partial recovery. Based on our structural model, the disincentive effect is substantially larger and can amount to 15 percentage-points in case of full recovery. Second, disincentive effects may be less relevant for recovery than at the onset of disability as employment effects of recovery may be less strong than the effect of negative health shocks. Earlier studies estimated the effect of negative health shocks to range from 7 to 15 percentage-points in terms of employment (García Gómez et al., 2013; Lindeboom et al., 2016). These estimates indeed exceed our implied employment response to mental health recovery of 4 percentage-points for those without benefits.

Since benefit disincentive effects to resume work after recovery are found to be substantial, there is a strong case for the continuous screening of the earnings capacity of DI benefit recipients. In the current system, re-examinations of DI benefit recipients are rare. Continuous screening of DI benefit recipients could identify those that have (partially) recovered, and benefits could be adjusted accordingly. Re-examinations may also stimulate re-employment rates by eliminating the ex ante disincentive effects of benefits.[36]

[36]Changes to the DI benefit setup could also increase work resumption rates upon recovery. We explore potential changes through the structural model, as shown in Online Appendix B.4. These

References

Autor, D., & Duggan, M. G. (2006). The growth in the social security disability rolls: a fiscal crisis unfolding. *Journal of Economic Perspectives*, *20*(3), 71–96.

Borghans, L., Gielen, A. C., & Luttmer, E. F. (2014). Social support substitution and the earnings rebound: Evidence from a regression discontinuity in disability insurance reform. *American Economic Journal: Economic Policy*, *6*(4), 34–70.

Bound, J. (1989). The health and earnings of rejected disability insurance applicants. *The American Economic Review*, *79*(3), 482–503.

Browning, M., Moller Dano, A., & Heinesen, E. (2006). Job displacement and stress-related health outcomes. *Health Economics*, *15*(10), 1061–1075.

Campolieti, M., & Riddell, C. (2012). Disability policy and the labor market: evidence from a natural experiment in Canada, 1998–2006. *Journal of Public Economics*, *96*(3-4), 306–316.

CBS. (2019). *Sociale zekerheid; kerncijfers, uitkeringen naar uitkeringssoort.* CBS Statline.

Chen, S., & Van der Klaauw, W. (2008). The work disincentive effects of the disability insurance program in the 1990s. *Journal of Econometrics*, *142*(2), 757–784.

Claussen, B., Bjørndal, A., & Hjort, P. F. (1993). Health and re-employment in a two year follow up of long term unemployed. *Journal of Epidemiology & Community Health*, *47*(1), 14–18.

Dano, A. M. (2005). Road injuries and long-run effects on income and employment. *Health Economics*, *14*(9), 955–970.

French, E., & Song, J. (2014). The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy*, *6*(2), 291–337.

García Gómez, P., & López Nicolás, Á. (2006). Health shocks, employment and income in the Spanish labour market. *Health Economics*, *15*(9), 997–1009.

García Gómez, P., Van Kippersluis, H., O'Donnell, O., & Van Doorslaer, E. (2013). Long-term and spillover effects of health shocks on employment and income. *Journal of Human Resources*, *48*(4), 873–909.

Garcia Mandico, S., García Gómez, P., Gielen, A., & O'Donnell, O. (2018). Earnings responses to disability benefit cuts. *IZA Discussion Paper*, *No. 11410*.

changes could however also influence the inflow rates into DI.

Godard, M., Koning, P., & Lindeboom, M. (2019). Targeting disability insurance applications with screening. *IZA Discussion Paper*, *No. 12343*.

Gruber, J. (2000). Disability insurance benefits and labor supply. *Journal of Political Economy*, *108*(6), 1162–1183.

Hausman, C., & Rapson, D. S. (2018). Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics*, *10*, 533–552.

Koning, P. (2019). Verleden, heden en toekomst van de arbeidsongeschiktheid. *Economische Statistische Berichten*.

Koning, P., & Lindeboom, M. (2015). The rise and fall of disability insurance enrollment in the Netherlands. *Journal of Economic Perspectives*, *29*(2), 151–72.

Koning, P., & van Sonsbeek, J.-M. (2017). Making disability work? The effects of financial incentives on partially disabled workers. *Labour Economics*, *47*, 202–215.

Korpi, T. (2001). Accumulating disadvantage. Longitudinal analyses of unemployment and physical health in representative samples of the Swedish population. *European Sociological Review*, *17*(3), 255–273.

Kostol, A. R., & Mogstad, M. (2014). How financial incentives induce disability insurance recipients to return to work. *American Economic Review*, *104*(2), 624–55.

Kuhn, A., Lalive, R., & Zweimüller, J. (2009). The public health costs of job loss. *Journal of Health Economics*, *28*(6), 1099–1115.

Lechner, M., et al. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends in Econometrics*, *4*(3), 165–224.

Lindeboom, M., Llena-Nozal, A., & van der Klaauw, B. (2016). Health shocks, disability and work. *Labour Economics*, *43*, 186–200.

Lindelow, M., & Wagstaff, A. (2005). *Health shocks in China : are the poor and uninsured less protected ?* (Policy Research Working Paper Series No. 3740). The World Bank.

Low, H., & Pistaferri, L. (2015). Disability insurance and the dynamics of the incentive insurance trade-off. *American Economic Review*, *105*(10), 2986–3029.

Maestas, N. (2019). Identifying work capacity and promoting work: A strategy for modernizing the SSDI program. *The ANNALS of the American Academy of*

Political and Social Science, *686*(1), 93-120.

Maestas, N., Mullen, K. J., & Strand, A. (2013). Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt. *The American Economic Review*, *103*(5), 1797–1829.

Moore, T. J. (2015). The employment effects of terminating disability benefits. *Journal of Public Economics*, *124*, 30–43.

Morris, J. K., & Cook, D. (1991). A critical review of the effect of factory closures on health. *Occupational and Environmental Medicine*, *48*(1), 1–8.

OECD. (2019). *Social expenditure database.*

Paul, K. I., & Moser, K. (2009). Unemployment impairs mental health: Meta-analyses. *Journal of Vocational behavior*, *74*(3), 264–282.

Pettersson-Lidbom, P., & Thoursie, P. S. (2013). Temporary disability insurance and labor supply: evidence from a natural experiment. *The Scandinavian Journal of Economics*, *115*(2), 485–507.

Portaal, A. (2019). *Wet verbetering poortwachter.* (data retrieved from Arbo Portaal, `https://www.arboportaal.nl/onderwerpen/wet-verbetering-poortwachter`)

Salm, M. (2009). Does job loss cause ill health? *Health Economics*, *18*(9), 1075–1089.

Schmitz, H. (2011). Why are the unemployed in worse health? The causal effect of unemployment on health. *Labour Economics*, *18*(1), 71–78.

Staubli, S. (2011). The impact of stricter criteria for disability insurance on labor force participation. *Journal of Public Economics*, *95*(9-10), 1223–1235.

Sullivan, D., & Von Wachter, T. (2009). Job displacement and mortality: An analysis using administrative data. *The Quarterly Journal of Economics*, *124*(3), 1265–1306.

UWV. (2012). *Statistische tijdreeksen 2012.* (data retrieved from UWV, `https://www.uwv.nl/overuwv/kennis-cijfers-en-onderzoek/statistische-informatie/statistische-tijdreeksen-2012.aspx`)

UWV. (2018). *Statistische tijdreeksen 2018.* (data retrieved from UWV, `https://www.uwv.nl/overuwv/Images/uwv-tijdreeksen-2018.pdf`)

UWV. (2019a). *Kwantitieve informatie 2018.* (data retrieved from UWV, `https://www.uwv.nl/overuwv/Images/uwv-kwantitatieve-informatie-2018.pdf`)

UWV. (2019b). *Wia uitkering.* (data retrieved from UWV, https://www.uwv.nl/particulieren/ziek/ziek-wia-uitkering/tijdens-wia-uitkering/detail/hoelang-krijg-ik-wga/hoelang-krijg-ik-loongerelateerde-uitkering-lgu)

Van Sonsbeek, J.-M., & Gradus, R. H. (2012). Estimating the effects of recent disability reforms in the Netherlands. *Oxford Economic Papers*, *65*(4), 832–855.

Von Wachter, T., Song, J., & Manchester, J. (2011). Trends in employment and earnings of allowed and rejected applicants to the social security disability insurance program. *The American Economic Review*, *101*(7), 3308–29.

Weathers, R. R., & Hemmeter, J. (2011). The impact of changing financial work incentives on the earnings of social security disability insurance (SSDI) beneficiaries. *Journal of Policy Analysis and Management*, *30*(4), 708–728.

A Appendix

A.1 Mental health care expenditures and non-mental health care expenditures

Table A.1: Construction of mental healthcare expenditures and physical healthcare expenditures based on expenditure categories used by Statistics Netherlands

Expenditure category[a]	Mental healthcare	Physical healthcare
General practitioner		X
Pharmacie		
Dental healthcare		
Hospital healthcare		X
Paramedical healthcare		X
Apparatus		
Hospital transportation		
Birth care		
Health care expenditures incurred abroad		
Other cost		
First-line psychological healthcare	X	
Mental healthcare	X	
Basic-mental healthcare	X	
Specialist mental healthcare	X	
Geriatric rehabilitation healthcare	X	
Nursing without stay		X
Sensory disability healthcare		

Note: (a) Expenditure categories as used by Statistics Netherlands

A.2 Additional empirical results

Figure A.1: Income sources relative to application for the selected sample relative to the moment of application

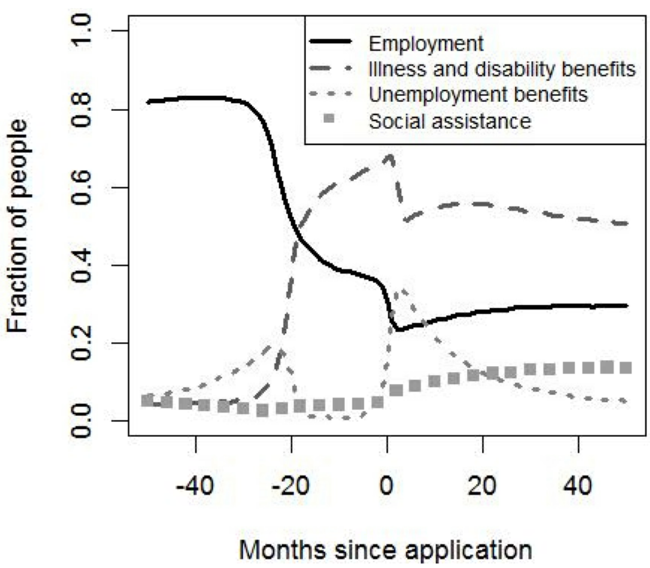


Table A.2: Descriptive statistics by broad degree-of-disability categories

	Degree of disability:		
	0	0-35	35-80
Age	45.764	46.966	49.480
Female	0.578	0.559	0.505
Dutch native	0.615	0.629	0.680
Education:			
Unknown	0.154	0.076	0.149
Low	0.364	0.309	0.228
Middle	0.370	0.444	0.355
High	0.112	0.171	0.268
Employment:			
Pre-application hourly wage	10.243	14.531	18.948
Pre-application hours	29.917	32.497	35.118
Mental health:			
Treatment length^b	34.371	32.491	34.567
DI application:			
FML	8.623	10.222	11.752
Disability percentage	0.000	19.251	54.449
Post-application hourly wage^c	11.686	12.008	11.757
Post-application hours^d	30.689	31.857	24.164
Chance of health improvement^e:			
NA	0.169	0.284	0.249
Reasonable to good	0.734	0.636	0.693
Small	0.077	0.077	0.055
Non-existent	0.020	0.004	0.003
Medical expenditures:			
Mental healthcare expenditures^f	6022	2861	3543
Physical healthcare expenditures^f	2195	1521	1685
Observations	9,502	9,498	6,080

Notes: ^{a}P-value of two-sample t-test for equality of means; ^{b}Length of the mental health treatment in months; ^{c}Potential post-application hourly wage as determined by UWV; ^{d}Potential post-application weekly working hours as determined by UWV; ^{e}Estimated chance of health improvement as determined by UWV; ^{f}Annual costs in euros in the year of DI application.

Figure A.2: McCrary discontinuity test of application densities at the 35% threshold of disability degrees

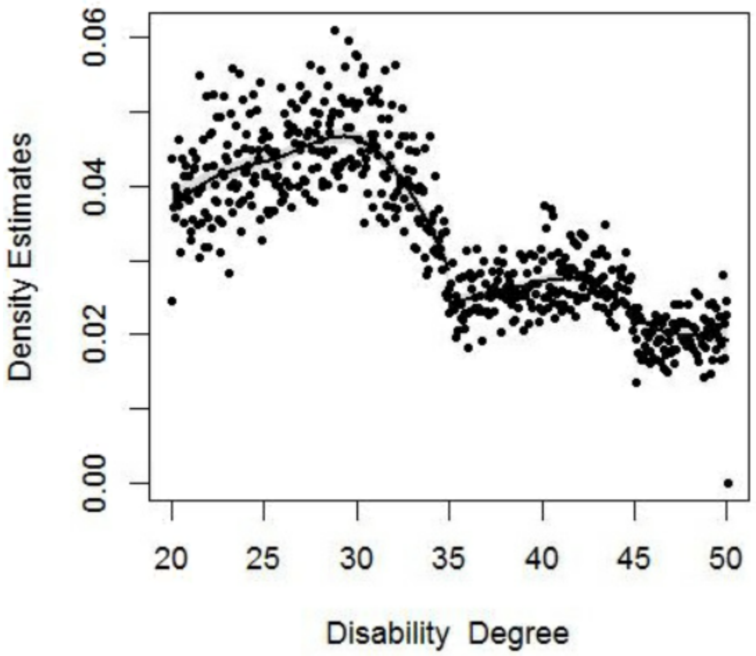


Table A.3: Additional robustness specifications on the unadjusted[a] DiD for employment, hours and hourly wage

	Specification				Outcome measures		
	Window[b]	Donut	N_c^c	N_t^d	Employment	Hours	Wage[e]
Baseline model	12	0	3,346	1,656	-0.035**	-4.984 **	0.225
					(0.002)	(0.207)	(0.252)
No covariates	12	0	3,346	1,656	-0.034**	-4.820**	0.005
					(0.002)	(0.208)	(0.261)
Monthly dummies	12	0	3,346	1,656	-0.035**	-4.984**	0.224
					(0.002)	(0.207)	(0.253)
Mean level	12	0	1	1	-0.034**	-4.820**	0.035
					(0.003)	(0.441)	(0.503)
Non-parametric	12	0	3,346	1,656	-0.034**	-4.820**	0.010
					(0.002)	(0.208)	(0.262)

Note: [a]Comparison between 20-30% and 40-50% disability degree groups; [b]Incorporated number of months before and after recovery; [c]Number of individuals in the control group; [d]Number of individuals in the treatment group; [e]Hourly wage; standard errors in parentheses; * significant at a 5% significance level; **significant after applying the $\frac{1}{60}$ Bonferroni correction factor

Table A.4: Unadjusted[a] DiD estimates based on various drops in mental healthcare expenditures

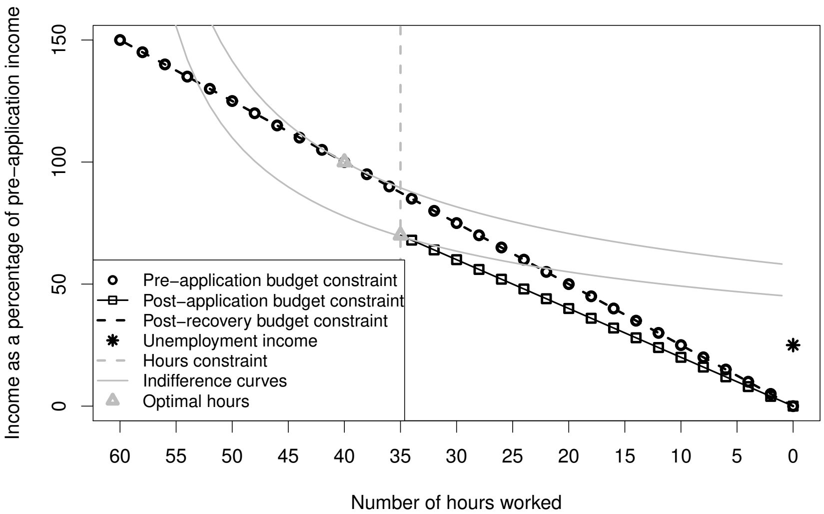
	Specification				Outcome measures
Decrease in healthcare cost	Window[b]	Donut	N_c^c	N_t^d	Employment
90 %	24	0	4,545	2,529	-0.069**
					(0.001)
80 % (Baseline)	24	0	4,735	2,683	-0.065**
					(0.001)
60 %	24	0	5,051	2,915	-0.061**
					(0.001)
40 %	24	0	5,308	3,140	-0.061**
					(0.001)

Note: [a]Comparison between 20-30% and 40-50% disability degree groups; [b]Incorporated number of months before and after recovery; [c]Number of individuals in the control group; [d]Number of individuals in the treatment group; standard errors in parentheses; * significant at a 5% significance level; **significant after applying the $\frac{1}{60}$ Bonferroni correction factor

A.3 Graphical illustration without DI benefits

To illustrate the labor supply decision for rejected DI applicants, we consider here an individual with an assessed degree of disability of 30%. The degree of disability comprises of an hours restriction of 87.5% and an hourly wage restriction of 80%. Figure A.3 shows the budget constraints for this individual, together with utility indifference curves.

Figure A.3: Budget constraints of a fictitious individual who has not been awarded disability benefits



The pre-application budget constraint, shown by the dotted line, is a straight line in which income increases one-to-one with wage earnings. The only non-linearity in the pre-application budget constraint is caused by the presence of social assistance. The post-application budget constraint is shown by the line with squares. The slope of the budget constraint is lower due to a reduction in hourly wage of 20%, potentially causing the optimal number of working hours (shown by the gray triangles) to be reduced. Furthermore, the hours constraint, shown by the gray dashed line, is enforced in the post-application stage. If the optimal post-application number of working hours exceeds the hours constraint, the individual will either work the maximum number of hours possible (35 in this case) or not work at all. The post-recovery budget constraint, shown by the black dashed line, is identical to the pre-application budget constraint. The response to recovery is thus equal to the reverse response to the onset of the disability.

B Online Appendix

B.1 The pre-application process

The pre-application process differs based on the employment status of individuals at the moment their disability benefit starts. The following subsections discuss the pre-application process for those individuals who are employed when falling ill, and for individuals who are unemployed when falling ill.

The employed

If an employed individual falls ill, he/she must go through the following steps (Portaal, 2019):

1. Within one week, the illness has to be reported to the medical officer of the company/occupational health and safety service (referred to as medical officer in the remainder of the paper).
2. After six weeks of absence, the medical officer must make an analysis. This analysis must contain the reason the employer cannot work, the possibilities of recovery and the anticipated time of recovery.
3. Within eight weeks the employee and employer decide on a plan of action. The plan of action contains the activities both employee and employer will undertake to ensure that the employee can return to work.
4. If the illness continues, the employer must keep a reintegration plan listing all activities taken by employee and employer to get the employee back to work.
5. Once every six weeks, the employer and employee have to discuss the progress made. The employer assigns a case-manager to the case. The case-manager supervises and controls the implementation of the plan of action.
6. After 42 weeks, the illness of the employee is registered at the UWV, the employee insurance agency.
7. After 46 to 52 weeks the employee and employer evaluate the first year of sickness and determine which reintegration result they would like to achieve in the second year.
8. After 20 months, the employee and employer write a reintegration-report. This report lists all the actions taken so far, and their results.

9. If the employee is not able to return to work, he/she will receive a disability insurance application form in the 87th week.

10. If all requirements are met, the degree of disability of the employee is assessed by the UWV. If this degree is above 35%, disability benefits are awarded. The benefit payments start 2 years after the start of the illness.

Employers are obliged to pay 70% of the last paid earnings to the employee during illness. If the employer has not exerted sufficient effort to reintegrate the employee, the period in which the employer is obliged to pay 70% of the last paid earnings is extended by one year at maximum.

Unemployed

Next to employed workers, workers without a formal employer may be eligible for disability insurance. This group contains individuals whose (temporary) contract has ended and who receive unemployment benefits and on-call workers. For these individuals, the process up to the application moment is as follows:

1. In case someone is still employed at the moment he/she falls ill, the illness should be reported to the employer. The employer will report this to the UWV.
2. In case someone is no longer employed, the illness should be reported to UWV.
3. For the first 13 weeks of illness, unemployment benefits will be paid (if applicable). If unemployment benefits are not applicable, or if unemployment benefits are terminated during the first 13 week illness ,illness benefits will be paid immediately.
4. UWV assigns a re-integration-manager to the case.
5. The process continues in the same way as described for the employed person above, from point 2 onward, with UWV that replaces the employers' role.

B.2 Estimation details of the structural labor supply model

This section illustrates the estimation of the structural labor supply model. The model assumes utility maximization over the number of hours worked. We adopt a Cobb-Douglas utility function with utility weights normalized to one.[1] Since eligibility to

[1] Since the utility function is estimated on a single employment decision, we can identify one preference parameter at most.

social assistance depends on partner income, our focus is on income at the household level. The general utility maximization problem for individual i is as follows:

$$\max_{E_i} u(L_i, I_i) = L_i^{\lambda_i+\delta_g} I_i^{1-\lambda_i-\delta_g} \tag{B.1}$$

$$\text{s.t. } L_i = T - E_i \tag{B.2}$$

$$I_i(E_i) = E_i w_i + DI(E_i) + \tilde{I}_i \quad \text{if} \quad E_i w_i + DI(E_i) + \tilde{I}_i \geq SA_i \tag{B.3}$$

$$I_i(E_i) = (1-F)\ SA_i \quad \text{if} \quad E_i w_i + DI(E_i) + \tilde{I}_i < SA_i \tag{B.4}$$

with T the total amount of time an individual can divide between leisure L_i and employment E_i. We set $T = 60$, the highest observed pre-application number of working hours. I_i is total income, consisting of labor income ($E_i\, w_i$), potential DI benefits ($DI(E_i)$) and partner income ($\tilde{I}_i$). If income falls below the social assistance level, it is supplemented up to this level.[2] Income from social assistance is discounted by factor $1 - F$, resembling the stigma attached to receiving social assistance. Lastly, λ_i displays the individual specific utility parameter and δ_g is a group-specific shock to this utility parameter.

As explained in the main text, the model consist of three stages: the pre-application stage, the post-application stage and the post recovery stage. The following subsections discuss the estimation details of every stage in more detail.

Pre-application stage

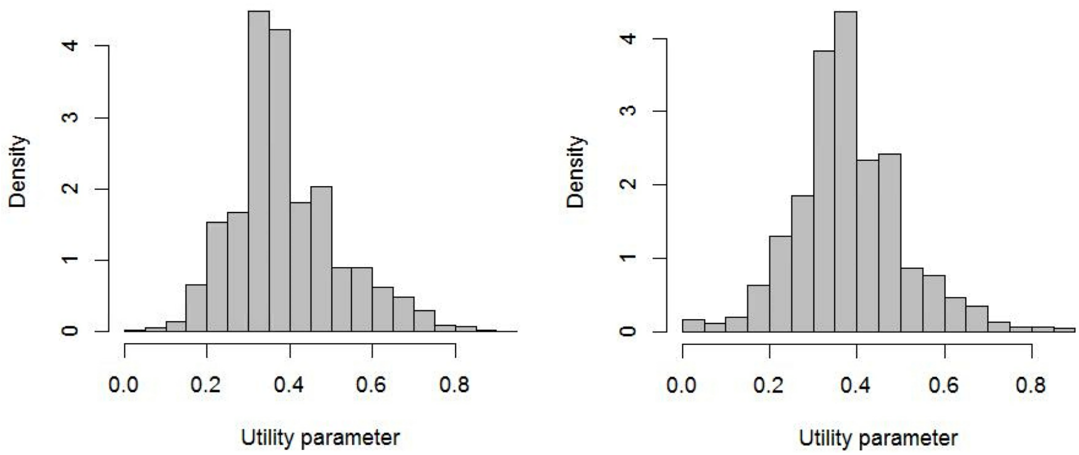
In the pre-application stage, all DI applicants work a positive number of hours and none receive disability benefits. Solving the individuals' pre-application utility maximization problem gives the following condition:

$$\lambda_i = \frac{(T - E_i)w_i}{\tilde{I}_i + T w_i} \tag{B.5}$$

Since all relevant variables are observed, we can determine λ_i for all workers in our sample. The resulting distributions of the utility parameters of workers that are awarded and rejected disability benefits later on are – given the similarities in pre-application employment histories – very similar (see Figure B.1).

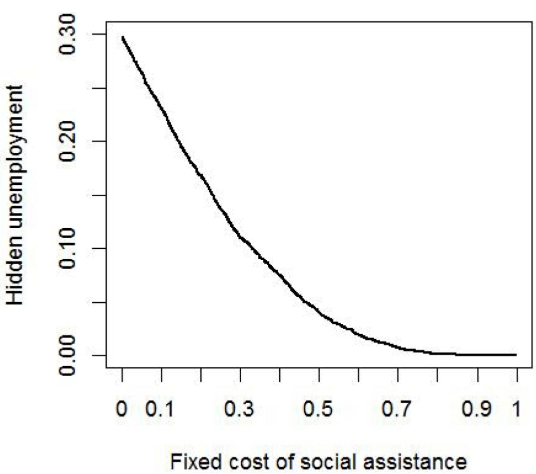
[2] The social assistance level is approximately €1,000 per month for singles and €1,500 for couples.

Figure B.1: Distribution of utility parameters for DI applicants with disabilities degrees of 20-30% (left) vs 40-50% (right)



For 29.5% and 30.5% of the individuals in the DI 20-30% and DI 40-50% groups, respectively, the utility of receiving social assistance exceeds the utility of working the observed number of working hours. As these percentages are substantial, a proportional fixed cost of receiving social assistance is included, corresponding to the stigma of receiving social assistance – see Equation (B.4). Capping this percentage at 5% implies a proportional fixed cost of not working of 0.463.[3] Accordingly, one euro received through social assistance is worth 0.537 euros received while working.

Figure B.2: Share of people preferring social assistance over working, as a function of the fixed cost of social assistance



[3]Using a maximum of 10% yields similar results, see Figure B.2.

Post-application stage

Given the individuals' pre-application utility function parameter λ_i, the optimal number of post-application working hours can be predicted and compared to the observed number of working hours. In doing so, our initial focus is solely on the effect of the assessed drop in hourly wages and in the possible impact of hours restrictions. This then yields outcomes of the optimal number of post-application working hours that are greatly overestimated (see Figure B.3). To improve the fit to the data, we therefore allow for a change in the utility parameter for individuals with and without disability benefits in our sample. Within these two groups, this change is assumed equal for all individuals. When setting the average prediction error equal to zero, the resulting changes in λ are similar for both groups (0.24 and 0.26 for those without and with DI benefits, respectively). The resulting prediction errors of the model incorporating the utility shocks are shown in Figure B.4

Figure B.4 shows that the prediction errors for those without benefits are centered around zero, whereas for the group with DI benefits there is bunching at approximately 20 hours. The model predicts that a substantial share of the individuals who have been granted disability benefits will choose to earn 50% of their remaining earnings capacity, corresponding to the discontinuity at which the disability benefits become linked to pre-application earnings. Given that most people are eligible for UI benefits during their application period, the short term incentives to earn 50% of the remaining earnings capacity are much smaller than the long term incentives, which might explain the over-prediction of working part-time. The incentive to earn at least 50% of one's remaining earnings capacity is not present for those without disability benefits.

Figure B.3: Prediction error in the post-application stage for DI applicants with disabilities degrees of 20-30% (left) vs 40-50% (right) in the absence of a utility shock

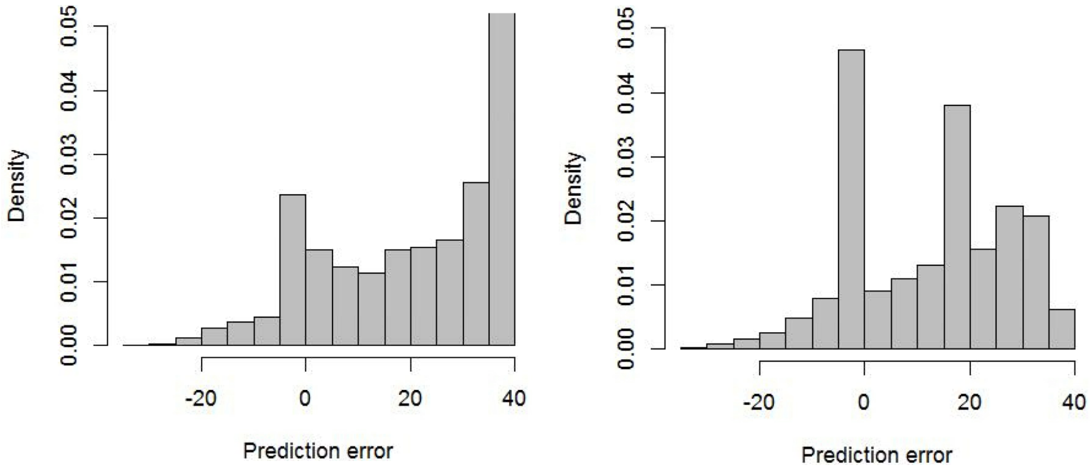
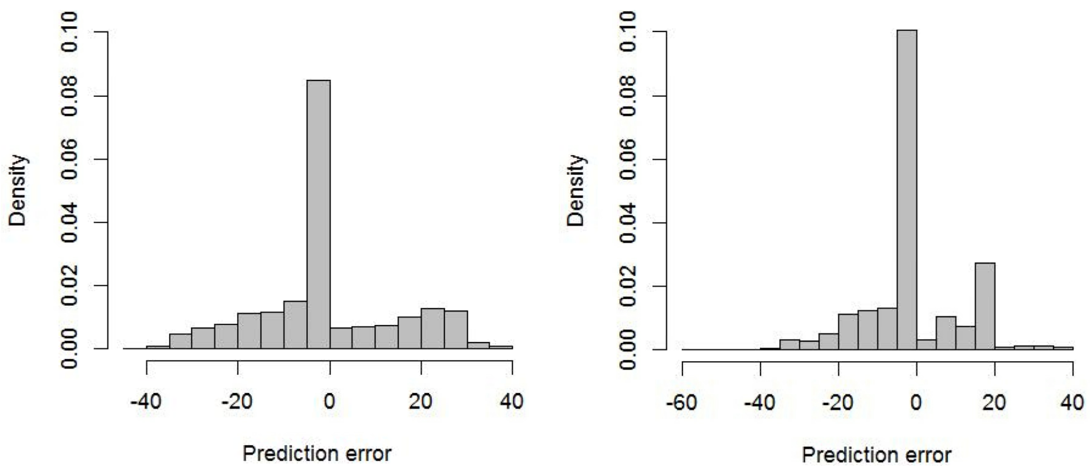


Figure B.4: Prediction errors in the post-application stage for DI applicants with disabilities degrees of 20-30% (left) vs 40-50% (right) incorporating group utility shocks



B.3 Simulated impact of recovery

The aim of the structural model is to assess the impact of full recovery for individuals with and without partial disability benefits. In the context of our model, recovery can be interpreted in two ways. First, it may be that recovery only implies a return of earnings capacity to the pre-disability level. However, with utility parameters not returning to their pre-application levels, the effect on working hours will be limited. Such persistence effects may originate from e.g. continued mental or physical strain from working, habit formation or increases in household production. Alternatively, there also could be "full" recovery in the sense that both the earnings capacity and the utility function are restored to their pre-application level. The resulting employment outcomes for both cases of recovery are shown in Table B.1.

Table B.1: Predicted employment rates and average number of working hours

	Pre-application	Post application	Recovery of earnings capacity	Recovery of earnings capacity and utility parameters
Employment				
DI 20-30%	94.2%	38.2%	57.3%	94.2%
DI 40-50%	96.6%	32.7%	50.8%	81.3%
Weekly working hours				
DI 20-30	31.9	7.2	11.2	31.9
DI 40-50	33.3	5.5	6.6	16.0

Recovery of earnings capacity causes an increase in the employment incidence of about 20 percentage-points for the groups with and without disability benefits. In terms of working hours, however, there is a difference in outcomes. Individuals without benefits (DI 20-30%) either do not respond at all or increase their number of working hours. For individuals with benefits (DI 40-50%), however, there are both increases and decreases in the optimal number of working hours, leading to a divergence in the number of working hours of both groups. Those who decrease their number of working hours do so to remain at either 50% or 100% of their remaining earnings capacity. Obviously, such cash-cliffs only exist for individuals with disability benefits.

If both the earnings capacity and the utility parameters return to their pre-application levels, the response to recovery is stronger. Those without DI benefits return to their pre-application employment status, which implies that 5% will remain unemployed (i.e., the fraction of hidden unemployment in the DI application inflow). Most of the individuals with DI benefits will not return to their pre-application employment status, due to the disincentives in the DI scheme. They choose to earn less than their pre-application earnings (clustering at 50% or 100% of their pre-recovery remaining earnings capacity) and a larger fraction of them chooses to remain unemployed.[4]

[4] See Figures B.5 and B.6 for the individual recovery responses.

Figure B.5: Recovery of earnings capacity: response for DI applicants with disabilities degrees of 20-30% (left) vs 40-50% (right)

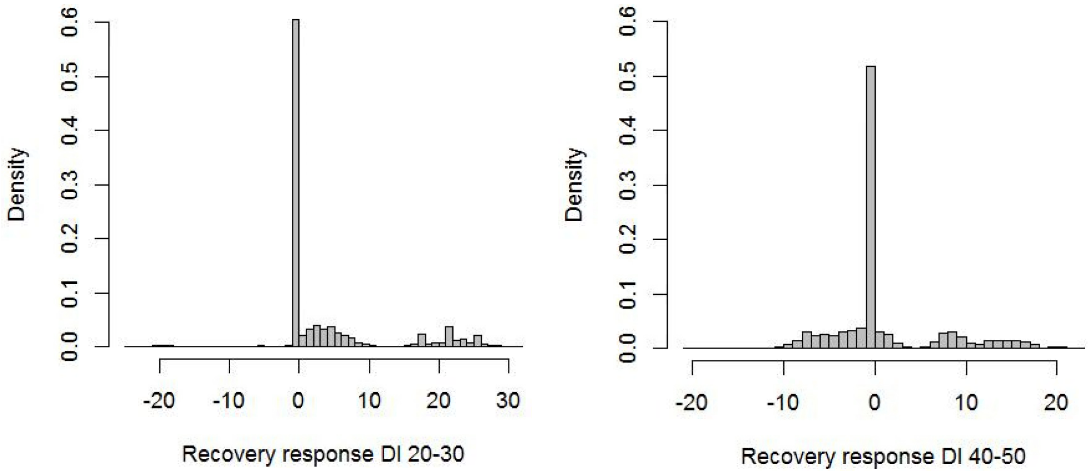
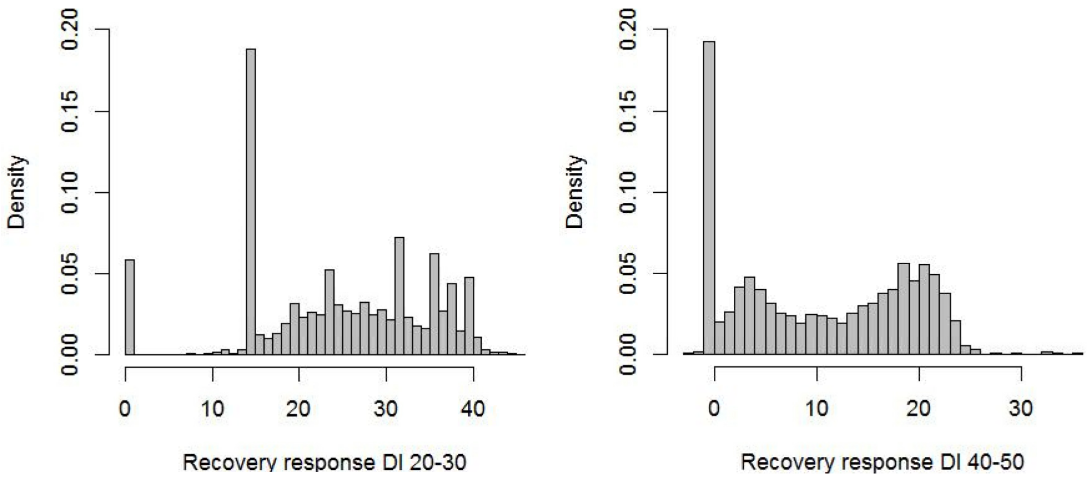
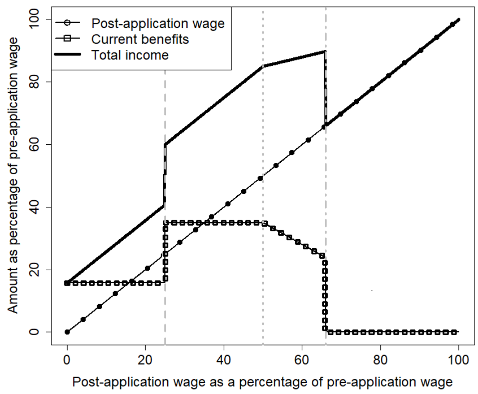


Figure B.6: Recovery of earnings capacity and utility function: response for DI applicants with disabilities degrees of 20-30% (left) vs 40-50% (right)

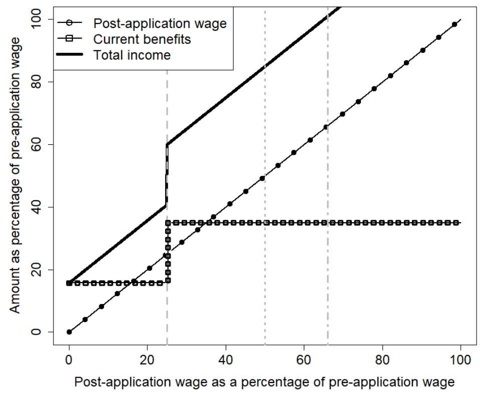


B.4 Simulation: policy analysis of DI benefit regimes

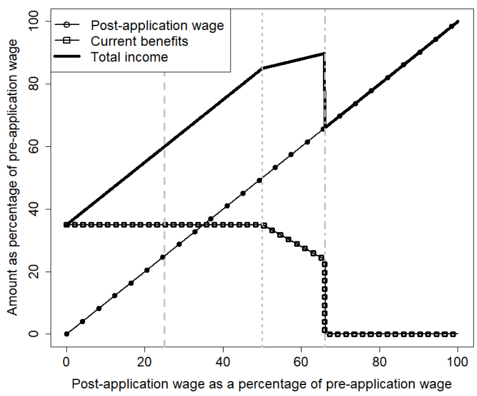
Figure B.7: Budget lines for alternative DI insurance regimes



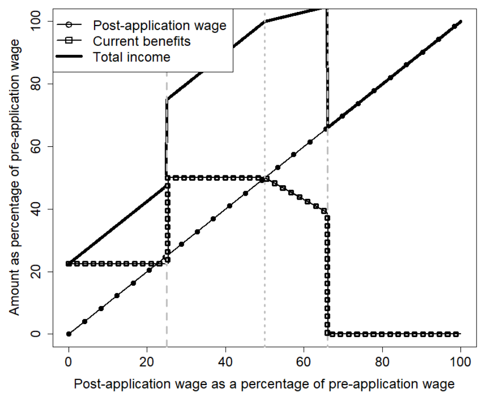
(a) Current DI system



(b) No earnings cap



(c) No minimum labor income requirement



(d) 100 % replacement rate

Given that we find disincentives of DI benefits upon recovery for the current Dutch system, a pertaining question for policy is whether changes in benefit conditions may limit these effects. To gain insight into this, we simulate the impact of changing some of the key parameters of the Dutch DI system. The three regime changes which we explore are: (1) The abolishment of earnings caps for DI receipt, (2) No minimum labor income requirement (for the wage subsidy) and (3) A replacement rate of 100%. The effects of these changes are demonstrated in Figure B.7. We show in Table B.2 that increasing the replacement rate to 100% is most effective in increasing the employment rate at every stage. Increasing the replacement rate increases the (positive) cash cliff at 50% of the remaining earnings capacity, thereby increasing the incentive to work.

Clearly, increasing the amount of the benefits could increase the inflow, making a cost-benefit analysis of the policy change infeasible given the available data.

Table B.2: Counterfactual labor supply of individuals with DI benefits with different benefit regimes

Employment rate	Post-application	Recovery of earnings capacity	Full recovery
Without DI benefits	31.0%	54.1%	96.6%
Current benefits	32.7%	50.8%	81.3%
No earnings cap	32.7%	50.8%	85.4%
No income requirement	28.1%	44.2%	81.3%
100% replacement rate	45.2%	61.7%	89.1%
Weekly working hours			
Without DI benefits	5.2	10.0	33.3
Current benefits	5.5	6.6	16.0
No earnings cap	5.5	6.6	26.5
No income requirement	2.8	5.6	16.0
100% replacement rate	7.4	7.0	17.1