Mental health, labor supply and disability insurance: does recovery matter?

Pierre Koning[1,2,3,4], Paul Muller[1,2], Roger Prudon[1,2, *]

November 14, 2019

Preliminary: Do not cite or distribute

Abstract

A large share of Disability Insurance recipients suffer from mental health problems and are expected to recover at some point. Contrasting expected health improvements, the outflow rates from these temporary disability schemes are negligible. This paper estimates the disincentive effects of disability benefits on the response to a mental health improvement using administrative data on all Dutch disability benefit applicants. Using the completion of mental health treatment as a proxy for health improvement/recovery, we estimate a difference-in-differences specification, comparing those below the DI threshold with those above. We find that disincentives substantially offset the response to health recovery: employment among benefit recipients increases with 1.5-2.5% points less than among those without benefits. The difference in the working hours response is 1.7-4.5 hours. Using drops in healthcare expenditures as an alternative proxy for recovery confirms these results, and shows that they extend to physical health improvements. We compare these results to predictions from a structural labor supply model, estimated using individual earnings potential from the DI application data. These predictions show that disincentives can be substantially larger in case of full recovery.

[*]Affiliations: [1] Vrije Universiteit Amsterdam, [2] Tinbergen Institute, [3] Leiden University, [4] IZA. Corresponding author: Roger Prudon (r.e.k.prudon@vu.nl). We thank Owen O'Donnell and Maarten Lindeboom 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). Under certain conditions, part of these microdata are accessible for statistical and scientific research. For further information: microdata@cbs.nl.

1 Introduction

Governments across the globe are altering their social security systems to make them more sustainable. One of the largest components of social security systems (in monetary amounts) in the western world 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. Throughout the Western world, 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. Even though a large share of those with mental health problems are expected to recover (Korpi, 2001), the work-resumption rate of those suffering from mental health problems is low (Claussen, Bjørndal, & Hjort, 1993).

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[1]. As one of the first studies that considers DI outflow, we investigate the role of (mental) health recovery in the return to employment for DI recipients. Comparing DI applicants below and above the insurance cutoff, we estimate the effect of receiving benefits on the labor supply response to a positive health shock (mental health recovery). Thus, we assess whether DI benefits create disincentives for returning to work once health conditions improve. Considering all DI applicants in the Netherlands since 2006, we show that health improvements indeed coincide with an increase in labor supply. Disincentives from receiving benefits are significant however, as those with DI benefits respond substantially less than those without.

Contrasting the idea that disability schemes should be temporary for many, outflow rates of DI benefit recipients are mostly negligible (David & Duggan, 2006). Yet, research on outflow rates is scarce and insights into health and recovery patterns of DI recipients and the effect of health changes on their employment status are lacking. Research on the effect of changes in financial incentives of DI applicants (Koning & van Sonsbeek, 2017; Kostol & Mogstad, 2014; Campolieti & Riddell, 2012; Weathers & Hemmeter, 2011) and on termination of DI benefits (Moore, 2015; Garcia Mandico, García Gómez, Gielen, & O'Donnell, 2018) have found significant work resumption rates. Whether or not the work resumption is due to improved health, has however not been investigated.

The main reason for the lack of research on the interplay of DI, 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 administrative data covering the entire population of DI applicants since 2006 (over 600,000 individuals). We combine (i) DI application records (including detailed assessment outcomes), (ii) monthly administrative records on employment and

[1] As expected, 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).

social insurance recipience and (iii) administrative records describing mental health treatment. Using the DI application data we are able to identify applicants just above and below the DI 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 an indication of full recovery, we interpret the end of a mental health treatment trajectory as signifying a substantial improvement in health. We compare labor supply responses around this end date for those that receive DI benefits with those that do not. This comparison identifies the disincentive effects of DI benefits on the return to employment after (partial) recovery.

Our approach constitutes a difference-in-differences estimator, as we compare those with and without DI benefits, before and after recovery. Using employment histories we assess 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 difference-in-differences estimator only requires the assumption that it proxies recovery equally well in the two groups. A similar point holds for the issue of reverse causality: in some cases it may be employment that *causes* health improvement[2]. Again, this is unproblematic as long as reverse causality is equally strong in both groups.

We observe that around the time of recovery the employment rates start to diverge, as those without DI benefits start working at a higher rate than those with DI benefits. Disincentives are substantial: we find a negative impact of DI benefits of 1.5% point 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. Using a second proxy for health improvement, based on significant drops in healthcare cost, we confirm our results and show that they even 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. In addition, we find similar results (with slightly larger estimates) when constructing an alternative measure of the timing of recovery based on individual healthcare expenditures.

To interpret the magnitude of our findings, we benchmark the estimates against predictions from a structural labor supply model. The model is estimated using pre-sickness labor supply decisions. The DI application decision data provides an estimate of remaining earnings capacity, which allows validating the model by comparing predicted labor supply after disability with actual labor supply after the DI application. Based on the calibrated model,we simulate labor supply responses to health recovery. Assuming full recovery (defined as earnings capacity returning to its pre-application level), we find large disincentive effects of DI benefits: the increase

[2]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 inconclusive. Some of these studies have found negative effects of job loss on health, whereas others have found no such effects. The same holds for mental health.

in employment after recovery is 16 %-points lower for those with benefits than for those without benefits. This large difference should be considered as an absolute upper bound. It assumes (i) full recovery to pre-disability earnings capacity, (ii) no frictions in finding employment after recovery and (iii) no changes in preferences over the process of disability and recovery. The fact that our empirical findings are substantially smaller is most likely due to the end of mental health treatment not coinciding with full recovery. For some, health treatment may end if health has improved partly. For others it may even end simply due to absence of any prospects of recovery.

Our findings contribute to two strands of literature; literature on the effects of financial incentives of DI schemes on employment and the broader literature on the interplay between health and labor supply. Disincentive effects of DI benefits on employment have been shown to exist in a variety of settings. First of all, DI beneficiaries are less likely to be employed than those whose DI application has been rejected (Bound, 1989; Von Wachter, Song, & Manchester, 2011; Maestas, Mullen, & Strand, 2013; French & Song, 2014; Chen & Van der Klaauw, 2008). Additionally, DI beneficiaries whose benefits are terminated have high work resumption rates. Moore (2015) exploits the exclusion of drug and alcohol addictions as qualifying conditions for the disability insurance system in the US whereas Garcia Mandico et al. (2018) evaluate a reassessment of disability benefit claimants in the Netherlands. Both studies find work resumption rates of approximately 20% points for those whose benefits were terminated. As reassessments are done some time after the initial assessment, they could be used to examine the effect of health recovery for benefit claimants. However, Moore (2015) and Garcia Mandico et al. (2018) both examine reassessments in which the eligibility criteria have been tightened. It is therefore not possible to disentangle the effect from potential health recovery and the effect from the tightened eligibility criteria. Due to the fact that large scale reassessments are costly, reassessments without stricter eligibility criteria are extremely rare. The results of these studies do however show that there might be disincentive effects of receiving disability benefits which prevent (recovered) beneficiaries from returning to the work force. We show that these disincentives indeed play a role. Lastly, the work resumption rate of DI beneficiaries depends on their financial incentives: if the financial incentives to return to the workforce increase, for example through a drop in the replacement rate, the work resumption rate increases(Koning & van Sonsbeek, 2017; Kostol & Mogstad, 2014; Campolieti & Riddell, 2012; Weathers & Hemmeter, 2011). The disincentive effects of DI benefits are thus evident. However, none of the above mentioned studies considers how health improvements interact with these disincentives.

The second strand of related literature examines the relationship between health and labor supply. This literature can be divided into research on the effect of (negative) health shocks on labor supply on the one hand and the effect of employment on health on the other hand. 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 negative causal relationship has been established between negative health shocks and labor supply. The employment rate drops by 5 to 7% points due to a negative health shock.[3] The causal impact of health improvements, as considered in this paper, has however not been examined yet due to a lack of 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 (mental) health improvements instead of 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 illustrates the structural labor supply model and section 6 concludes.

2 Disability insurance in the Netherlands

The focus of this paper is on disability insurance in the Netherlands. The disability insurance system is managed by the employee insurance agency (UWV). This section will highlight the most important aspects of this system, and the (dis)incentives it creates for both (potential) applicants and for those receiving benefits. The first subsection discusses the historical background whereas the subsequent subsections explain the application process and the incentives they face. For a more detailed description, we refer to Koning and Lindeboom (2015).

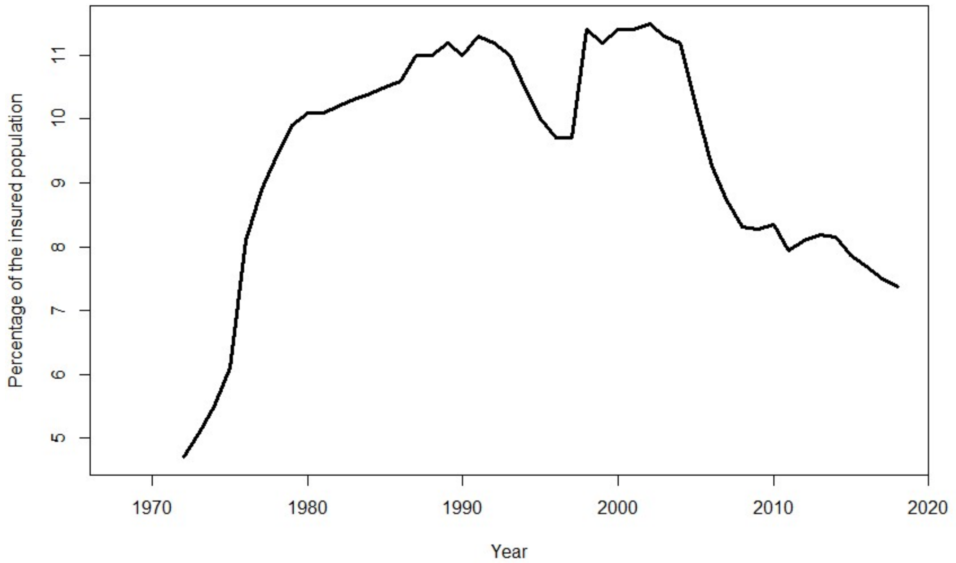
2.1 Historical background

The proportion of people receiving disability insurance 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 (which will be discussed in more detail below) was to replace the old disability benefit system (WAO) with a new disability benefit system (WIA). The new system was comprised of a benefit system for partially and temporarily disabled individuals (WGA) and a system for fully and permanently disabled individuals (IVA). As can be seen from Figure 2, the successive reforms led to a strong reduction in the number of new recipients 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

[3] These estimates refer to the effects of acute unscheduled hospitalizations and road injuries. The health shocks considered are thus very heterogeneous in magnitude. The magnitude of the effect based on self-assessed health shocks is difficult to interpret given the subjective scale of these shocks.

Figure 1: Percentage of the insured population receiving disability benefits in the Netherlands



Source: UWV (2012, 2018)

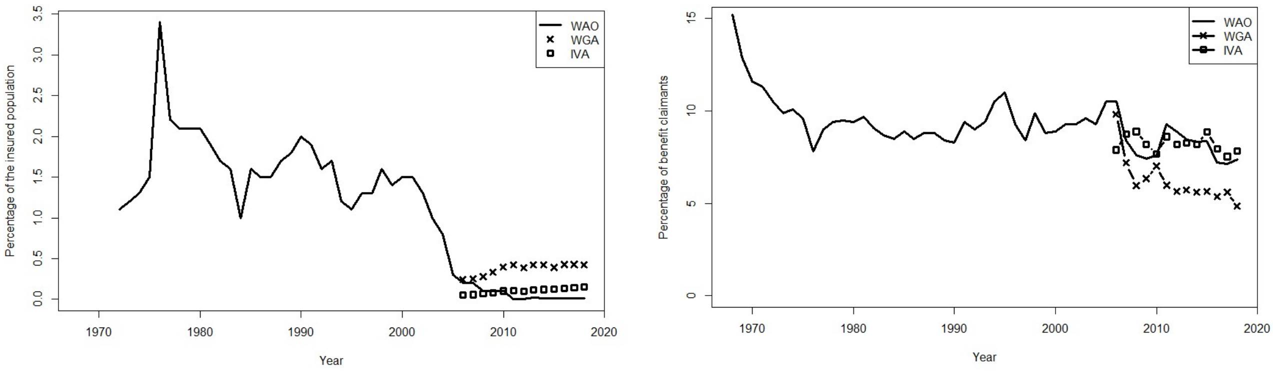
increased the responsibility of employer and employee in the waiting period leading up to the application. It specifies actions to be taken by both parties to accommodate the reintegration of the sick employee. The implementation of the Gatekeeper Protocol is estimated to reduce inflow into disability insurance by 30 to 40 % (Koning & Lindeboom, 2015; Van Sonsbeek & Gradus, 2012).

In 2006 several policy changes were implemented. First, stricter assessments for new DI applicants were imposed. For these applicants, the range of possible "reference" jobs was widened, potentially lowering the assessed degree of disability. This resulted in a higher rejection rate and lower disability benefits. A second change concerned the waiting period, a period in which the employer had to continue paying wages to a sick employee before the employee could apply for disability benefits. The waiting period was extended from one year 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 disability benefit recipients to use their remaining earnings capacity. The non-usage of remaining earnings capacity results in severe cuts in the disability benefits, as will be explained in more detail later on. These policy changes further reduced the inflow into the disability insurance system (Koning & van Sonsbeek, 2017).

2.2 Current disability insurance process

This paper considers all disability insurance applicants in the Netherlands between 2006 and 2017. In this period no major reforms were implemented. The disability insurance process from the start of the illness until the actual application depends on whether or not someone has a permanent employment contract. Both processes are described in detail in Appendix A.1. If

Figure 2: Inflow (left) and outflow (right) rates from various disability benefit systems



Source: UWV (2012, 2018)

someone has been ill for two years, he/she can apply for disability benefits. From the assessed earnings capacity and the pre-application earnings, the disability degree is derived as follows:

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

Individuals are than assigned to one of the following disability intervals: 0-35%, 35-45%, 45-55%, 55-65%, 65-80% and 80-100%. Benefit calculations are done based on the middle value of the assigned disability interval whereas the upper bound of the interval is used when determining the remaining earnings capacity. As an illustration, consider an individual with a pre-application wage of 3,000 euros per month. If it is assessed that this person can still work 16 hours (40 percent of full-time) and perform a job earning 2000 euros per month on a full-time basis, the remaining earnings capacity is set at $0.4 \times 2000 = 800$. In this case the degree of disability is $(1 - (800/3000)) * 100\% = 73.3\%$, implying that the disability interval is 65-80%.

If awarded benefits, benefit conditions differ between the so called wage related period and the continuation period (UWV, 2019b). The wage related period applies to people who in the 36 weeks prior to falling ill worked for at least 26 weeks. The benefits amount to 70% of the pre-application wage minus 70% of the wage currently earned. The duration of the wage related period is roughly one month for every year worked since the age of 18. The monetary amount and duration of the wage related period is equal to the unemployment benefit system.

Once the wage related period ends, the continuation period starts. The amount of benefits awarded depends on the utilization of the remaining earning capacity. If someone uses at least 50% of their remaining earning capacity, the amount of benefits remain linked to the pre-application wage. If someone earns 100% or more of the remaining earning capacity, benefits amount to 70% of the pre-application wage, minus 70% of the wage currently earned. If someone earns between 50% and 100% of the remaining earning capacity, benefits amount to 70% of the pre-application wage, minus 70% of the remaining earnings capacity. If someone uses less than

50% of their remaining earning capacity, benefits are no longer linked to the pre-application wage. Instead, benefits are based on the statutory minimum wage. The benefits are roughly 70% of the minimum wage, multiplied by the degree of disability. Individuals in the continuation period can no longer apply for unemployment benefits. They can however apply for social assistance, which amounts to approximately 70% of the minimum wage, i.e. similar to the disability benefits in case the remaining earnings capacity is not used.

The setup of the disability system is slightly different for individuals with an assessed degree of disability above 80%. These individuals are classified as fully disabled and their remaining earnings capacity is set at 0%. Their disability is assessed as being either temporary or permanent. If their disability is regarded as temporary, they receive DI benefits amounting to 70% of their pre-application earnings and they will be medically re-assessed after some time. If their disability is assessed as being permanent, they receive DI benefits amounting to 75% of their pre-application earnings and no re-assessments are performed.

Termination of benefits can occur for two different reasons. The first reason is that someone earns more than 65% of his/her pre-application wage. Earning more than 65% of the pre-application wage implies that the non-useable earnings capacity is below 35%, the threshold for getting benefits. The second reason for termination of benefit can occur if a re-evaluation by the UWV shows that the remaining earnings capacity exceeds 65% of the pre-application wage. The UWV thus believes that someone could earn more than 65% of the pre-application wage. Re-evaluations can be requested by benefit recipients themselves, by a former-employer and by the UWV itself. However, re-evaluations requested by disability applicants themselves, and re-evaluations based on earnings changes are given priority. Re-evaluations due to expected changes in health are therefore scarce (UWV, 2019a).

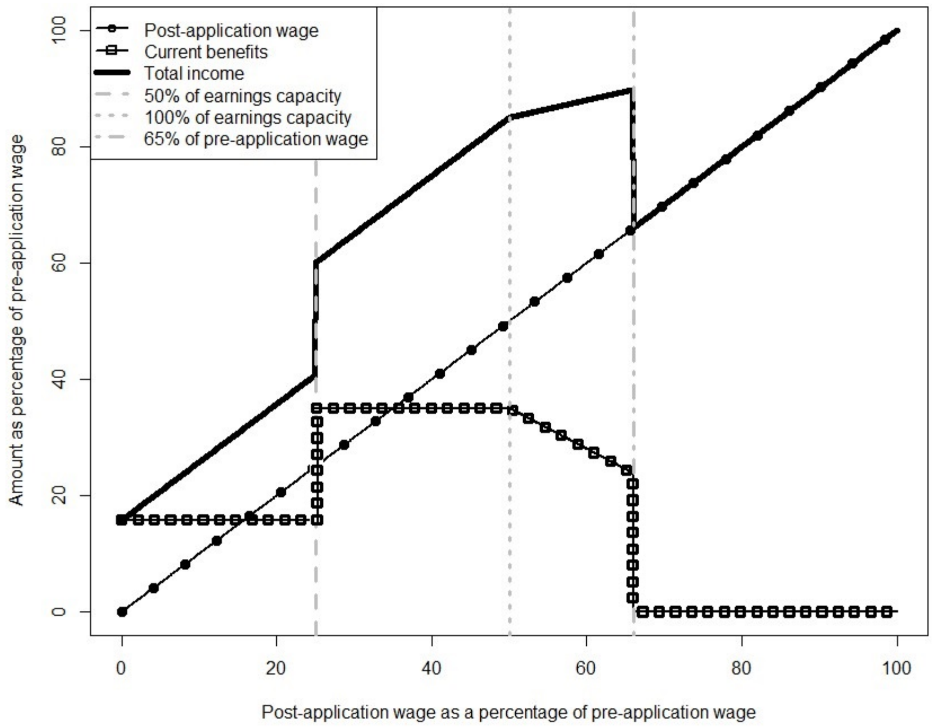
2.3 Economic incentives

This system creates economic (dis)incentives to work, that may differ before and after recovery. For illustrative purposes we show an example in which the remaining earnings capacity is 50% of pre-disability earnings, whereas the statutory minimum wage equals 50% of pre-disability earnings. Figure 3 shows the post-application wage, current benefits and total income as a function of the post-application wage. As the benefit system thresholds are based on percentages of pre-disability earnings, all amounts are shown as percentage of pre-disability earnings.

In Figure 3, the post-application wage is given by the line with dots. The line with squares represent the amount of benefits one receives, as his/her wage increases. The bold line represents total income, which is the sum of wage and benefits. Lastly, the dashed lines demarcate the relative earnings that are necessary to receive the wage subsidy (25%), the relative earnings at which the benefits starts to reduce(50%) as well as the maximum relative earnings that can be retrieved together with DI benefits (65%).

As long as earnings are below 50% of the remaining earning capacity, benefits are linked to the minimum wage and any increase in earnings does not affect the benefits. The increase in

Figure 3: Example total income as a function of post-application wage



earnings is not tapered by reductions in DI benefits. Once post-application wage exceeds 50% of the remaining earnings capacity, the benefits are linked to pre-disability wage and amount to of 70% of the pre-disability earnings, minus 70% of the remaining earning capacity. Since disability benefits are linked to the pre-disability wage instead of the minimum wage, there is a jump in the amount of benefits and thus in the amount of total income at 50% of the remaining earning capacity. This jump represents strong incentives to exploit more than 50% of the remaining earning capacity. Furthermore, the size of the jump is increasing in pre-application wage. The amount of benefits stays constant until the full remaining earning capacity is used. In the example of Figure 3, 50% of the remaining earning capacity equals 25% of the pre-application wage. The remaining earning capacity is 50% of the pre-application wage. At current income between 25% and 50%, benefits thus amount to 35% (70% minus 70% of 50%=35%).

If the individual earns more than 50% of his pre-disability earnings - thus exceeding the initial assessed earnings capacity - the earnings capacity will be adjusted accordingly and disability benefits will be lowered. This implies that an increase in the relative earnings with one % point goes together with a reduction of (relative) benefits of 0.7% point. Once the 65%-threshold is reached, all disability benefits are terminated. This creates a drop in total income at 65% of pre-disability earnings, resulting in strong disincentives to increase wage beyond 65% of pre-application wage. The magnitude of this drop is independent of the degree of disability. Income drops by 70% of 35% of pre-application income, i.e. 24.5% of pre-application income: total income drops 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 line in Figure 3. Total income equals, and thus always increases one-to-one with, current earnings. By comparing the total income of individuals with and without disability benefits, three main effects of having disability benefits emerge. First, there is an income effect as long as post-application wage is below 65% of pre-disability earnings. Individuals who receive disability benefits have higher total income due to the benefits. The expected income effect implies that they will most likely work less than they would have worked if they would not receive benefits. Second, there is an incentive effect for disability benefit recipients to earn at least 50% of the remaining earnings capacity. Lastly, there is a strong incentive to earn no more than 65% of pre-disability earnings, when receiving disability benefits.

Given these incentives, the response to recovery (defined as an increase in the potential hourly wage) of those with disability benefits is most likely 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, whereas those without disability benefits would return to full time employment, those with disability benefits have strong incentives not to do so. They would have to work full-time, instead of 65% of full time, to gain 10.5% points in income. This implies having to work one-and-a-halve times as much, to gain 10.5% points in income. The maximal response to recovery of an individual with benefits is thus most likely smaller than the maximal response of an individual without benefits. The incentive to earn at least 50% of the remaining earning capacity becomes slightly stronger, as less hours are needed to earn this amount. However, it is expected, and shown in the economic model in section 8, that the overall effect of the incentives induces individuals to respond less to recovery, than they would have done without disability benefits.

3 Data description

As the main interest of this paper lies in the labor response to recovery from mental health problems, for those with and without disability benefits, information on labor market behavior, disability applicants and (mental) health treatments were linked. Information on disability insurance applications was acquired from UWV and the earnings and mental health data were acquired from Statistics Netherlands (CBS). All data are administrative, and apply to all (relevant) Dutch citizens. The following subsections will discuss the various data sources and the selection of the sample.

3.1 Disability insurance application data

The disability insurance application data set is comprised of all applications between June 2006 and June 2017, which gives a total of 670,171 applications. One of the unique aspects of the data set is that it not only has information on individuals who have been awarded disability benefits, but also on individuals who have not been awarded disability benefits. The data set contains all information that is needed to determine the earnings capacity and the degree of disability. It includes the pre-application hourly wage and number of hours worked and the post-application potential hourly wage and number of working hours. 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. Whenever the analysis uses subgroups based on the degree of disability, those without a stated degree of disability will therefore be used as a separate group.

The application data includes the timing of the decision of the UWV on whether or not to award benefits. The decision could be made before or after the applicant is actually entitled to benefits, as benefits could be awarded retroactively or proactively. The timing of the decision is therefore a proxy of the actual implementation of the payment of benefits. As will be shown below, the proxy is accurate as benefit receipt changes strongly in the month of the decision.

Several health-related variables are included in the application information. The first group of variables concerns the medical diagnoses of the applicants. The diagnoses are either stated as so-called "CAS-codes" or as diagnoses groups created by the UWV itself. The CAS-codes are used by health and safety doctors responsible for the reintegration process of long-term absent workers of firms. They are comprised of a diagnose group letter, e.g. P for psychological diagnoses, and a three-digit number indicating the specific diagnose. The diagnoses variables used by the UWV itself contain 32 different groups of diagnoses. These are thus much broader than the CAS-codes. Applicants can have at most thee CAS-codes and three UWV diagnoses group codes. The CAS-codes are available for approximately 80% of all applications, whereas the UWV group codes are available for approximately 98% of all applications. The analysis will, where possible, combine both types of diagnoses information.

Besides the diagnoses information, the application information also contains information on the type and number of functional limitations of applicants, as assessed by a doctor of the 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 0, 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 the UWV. The probability of improvement can be assessed as "reasonable to good", "small" and "non-existent".

3.2 Income data

Income data is acquired from two separate CBS sources. The first source contains information on all employment contracts in the Netherlands between 2006 and 2018. From the income data, we extract information on an individuals' monthly earnings, hours worked and monthly employment status. In case an individual has multiple employment contracts at the same time, these employment contracts are combined to obtained the total wage and total number of hours worked. The employment indicator indicates whether an individual worked for at least one hour in a specific month. A panel of hourly wages is created by dividing the monthly wage by the number of hours worked per month. The monthly wage data does not discriminate between regular wage pay and bonus pay, causing outliers in the hourly wages. Approximately 200 out of the 2,000,000 hourly wage entries (0.01%) exceed 1,000 euros per hour. These outliers will be excluded from the analysis.

The second source of information contains indicators for various forms of income of all individuals in the Netherlands from 1999 till 2016. Using this information, a panel is created which indicates whether someone has a specific form of income in a specific month. The sources of income are: employment, self-employment, unemployment benefits, social assistance, old age pensions, disability or illness benefits and other social services. The disability or illness benefits indicator does not discriminate between disability and illness.

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, the employment panel obtained from the first source will be used.

3.3 Health data

Two separate sources of data on health treatments are used. The primary source, concerning data on mental health treatments, is derived from so called Diagnosis Treatment Trajectory (DBC) data. The DBC system is used in the Netherlands to simplify financing of medical treatments. Instead of having to pay for every single treatment, payment is based on treatment trajectories. These treatment trajectories comprise all treatments that are deemed necessary to alleviate or solve health problems. The mental health database used in this paper consists of all DBC's regarding mental health in the Netherlands between 2011 and 2016. The severity of the mental health problems of the sample considered is large, with average cost of treatment of approximately 5,500 euros, resembling roughly 140 hours of treatment. Every DBC entry states the starting date and ending date of treatment, even if this start or end is before or after the DBC period (in case treatment lasted for more than one year). For individuals with multiple treatment trajectories, the earliest start date and latest end date is used. This is done to ensure that the end of treatment actually implies that an individual no longer receives treatment. Information on whether an individual actually recovered because of the treatment is unfortunately not available. The end of treatment will therefore be used as a proxy for recovery.

The second source of health data contains the yearly healthcare expenditures as covered by basic health insurance. 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.2). Using the healthcare expenditures in the pre-application sickness period as a baseline measure, we create a proxy for recovery based on a drop in the healthcare expenditures. Several baseline measures and relative drops were use, yielding similar results. The final measure uses the year before the application date (as all DI applicants have been disabled during this entire year) as a baseline measure, and proxys the end of treatment year as the year in which healthcare expenditures drop below (and stay below) 20% of the baseline cost. Given that the expenditure data is only available on a yearly basis, there is more measurement error in the proxy compared to the proxy based on the DBC data.

3.4 Sample selection

Linking the various data sources discussed above yields a sample of disability insurance applicants for whom the application information, mental health information, and employment and income history is known. A selection was made to make the sample viable for analysis. This section will discuss the various selection steps.

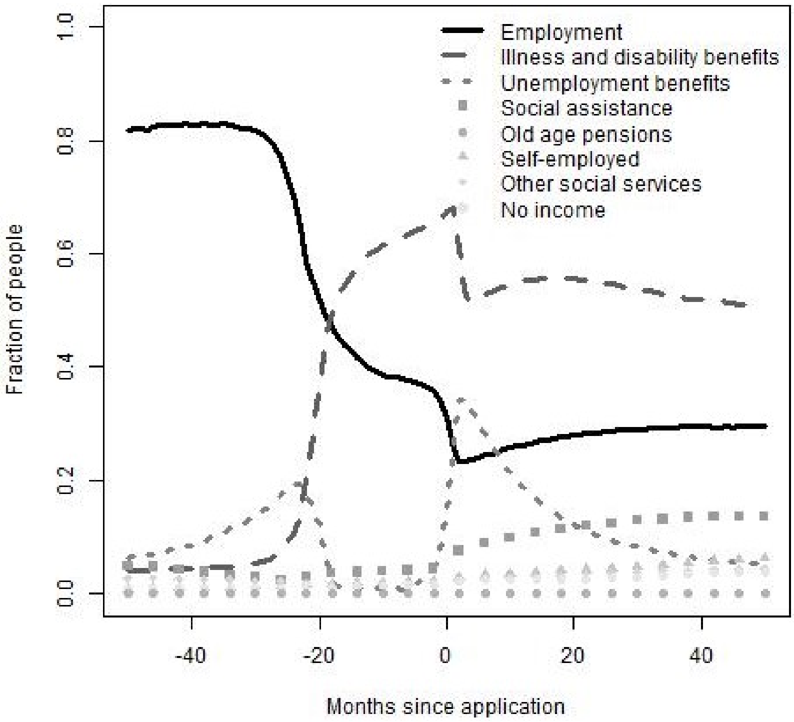
Table 1: Sample selection

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

Table 1 illustrates the various sample selection steps. The DI application data is comprised of 670,171 applications filed between January 2004 and July 2017. As re-examinations are treated as outcomes of labor and health changes, the sample is restricted to only include the first application of every individual. This leaves 605,757 individuals in the sample. Given that the employment data start in 2006, we only consider applications filed in 2006 or later, leaving 498,810 individuals in the sample. This ensures that the employment history of all included applicants is known at least from the moment of application. To ensure that the mental health problems are severe enough 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. Approximately 40% of the sample reported some mental health problem resulting in a sample of 191,056 individuals. As the analysis focuses on those who are expected to recover, we exclude the fully and permanently disabled, resulting in 116,403 temporarily and partially disabled individuals who applied due to some form of mental health problem in the time window considered.

We then link the application data to the mental health data. As the mental health data is only available until 2016, individuals are selected for whom the end of treatment occurred before the first of January 2016. This is done to ensure that a new mental health trajectory

Figure 4: Income sources relative to application for the selected sample

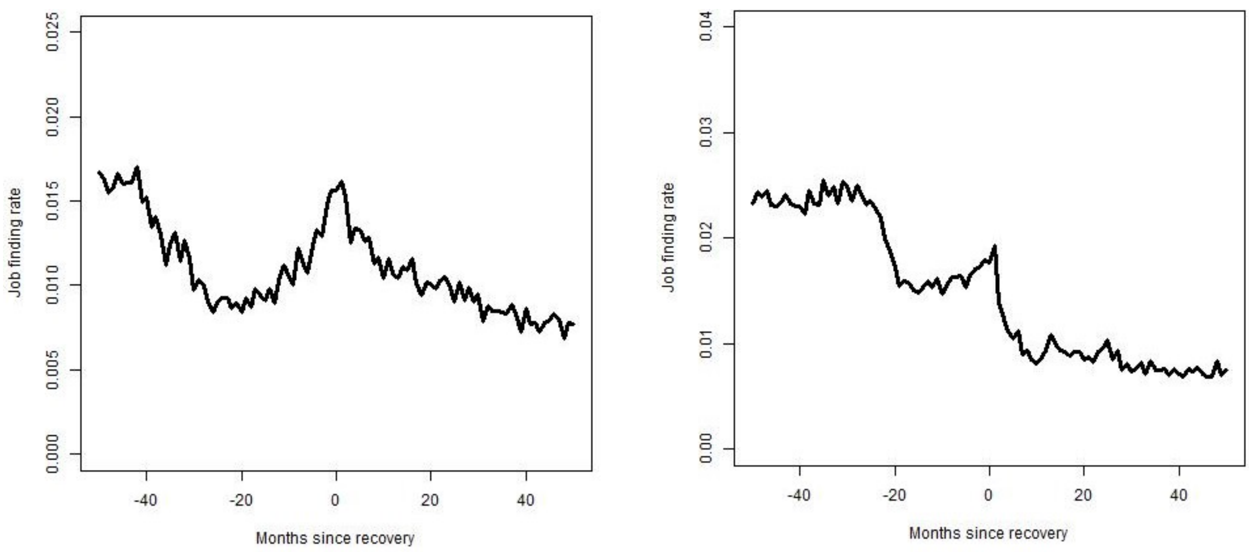


does not start closely after the observed end of treatment, e.g. an unobserved start of treatment in January 2017 while the end of treatment is defined as December 2016. The resulting sample consists of 45,966 individuals. The last selection step excludes all individuals for whom the end of treatment occurs before the application date. If the end of treatment would occur before the application date, it would be unclear whether an individual received disability benefits at the moment the treatment ended. The final sample consists of 25,080 DI applicants for who we have a proxy for mental health recovery.

Figure 4 shows the fraction of individuals in the sample receiving a certain form of income, relative to the application date. Note that the various sources of income are not exclusive, i.e. an individual can receive various sources of income simultaneously. The black line resembles the employment rate, which drops at the beginning of the 24 month waiting period[4]. Approximately half the sample continues to work during the waiting period. After the application, the employment rate drops further and stabilizes at approximately 30% points. The fraction of individuals receiving illness or disability benefits (the gray dashed line) mirrors the employment pattern. There is a sharp increase at the start of the waiting period, and a further small increase towards the application date. At application, the fraction of individuals receiving illness or disability benefits drops slightly, resembling those individuals whose application has been denied. After the application date, the fraction of individuals receiving disability benefits slowly decreases. The figure furthermore shows that there is a relatively large inflow from the unem-

[4]During the waiting period, we only observe whether people have a contract. We do not observe whether they actually work. Those identified as employed during the waiting period could thus in fact be staying at home ill. This is not possible after the waiting period.

Figure 5: Job finding rate (left) and job losing rate (right) relative to recovery



ployment benefit scheme into the DI scheme with about 20 percent of all individuals receiving unemployment benefits at the start of the waiting period. The spike in unemployment benefit recipients after the application date resembles individuals whose DI application has been denied. A second source of income for the rejected DI applicants is social assistance, which is received by approximately 10 percent of the sample after the application date. The remaining income sources are only relevant for a small portion of the sample and they do not show significant changes around the application date.

To assess whether the end of the mental health treatment does indeed proxy recovery, Figure 5 shows the hazard rates for finding and losing a job relative to 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. The end of mental health treatment does thus signal an improvement in the employment status for the selected sample. The hazard rate for job finding also shows that measurement error and/or anticipation seems to be present as the job finding rate already increases in the months leading up to recovery. The effect of measurement error and/or anticipation and how we deal with these phenomena will be discussed in the next section.

4 Difference-in-Difference (DiD)

To allow for causal inference on the disincentive effects of disability benefits on the employment response to mental health improvement, a DiD specification will be used. In a DiD framework, a control group is used to estimate the counterfactual for the treatment group (Lechner et al., 2011). Under the assumption that the control group and the treatment group follow parallel trends in the absence of treatment, any divergence between the groups can be attributed to the causal impact of the treatment. In our analysis, treatment is defined as recovery while receiving

DI benefits, whereas the control case is recovery while not receiving benefits. The resulting model specification is thus as follows:

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

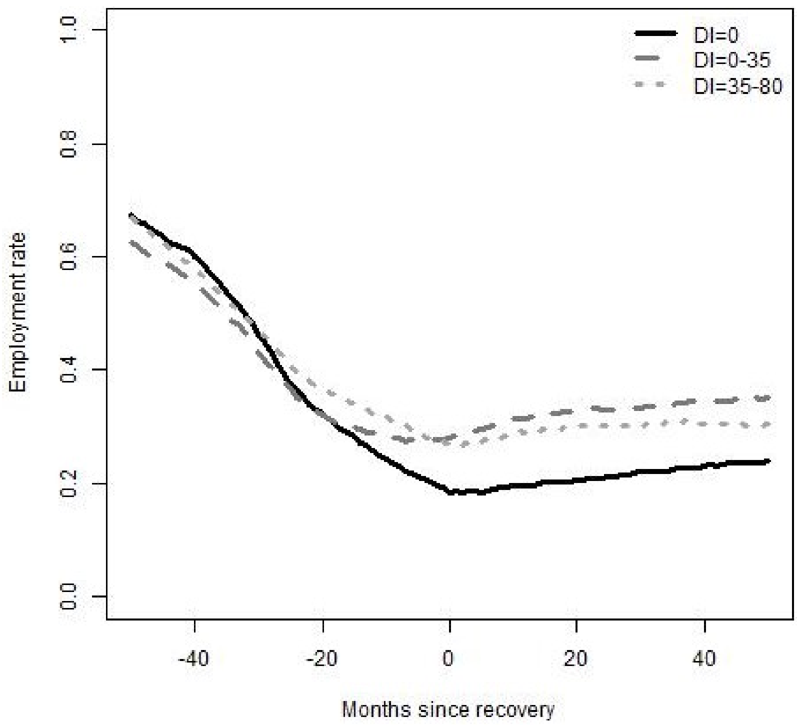
In which $\mathbb{1}_{DI_i}$ is an indicator function for receiving disability benefits and R_{it} is the period in which mental health treatment ends. This specification allows for fixed pre-recovery level differences between the two groups (α_1 and α_2) and flexible time trends $k(t)$ (we will use 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 addition to using employment as the outcome measure, we will also estimate (2) on hours worked and hourly wage. Given the panel nature of the data, there is a time component in the data which most likely exhibits high persistence. Therefore we cluster standard errors by individual (Hausman & Rapson, 2018). As a robustness test, two additional methods will be used to account for the possible presence of serial dependence: the model will also be estimated on mean levels and the analyses will be done fully non-parametric by not including any control variables in the regressions and thereby comparing the differences in unconditional means. This should circumvent any time-series characteristic issues such as heterogeneity (Lechner et al., 2011). Given the graphical evidence which will be shown below, a time window of 12 months before and 12 months after recovery will be used in the baseline models. Both the pre-recovery and post-recovery time trends are assumed to be linear, and the models control for age, gender, nationality, education and month and year dummies. As a robustness test monthly dummies (i.e. one month before recovery, two months before recovery, etc.) will be used instead of a linear time trend. Monthly dummies are used instead of for example a polynomial in time, as the dummies can accommodate any time trend.

Using a proxy for health affects the interpretation of the obtained estimates. The proxy implies that in both the treatment and control group not all individuals recovered after the end of treatment. If the proportion of individuals who recover is similar in the treatment and control group, the effects found in the DiD estimation are the result of the difference between the individuals who recover. Those who do not recover, both in the treatment and control group, will continue following the same trend after the end of treatment. If a share equal to ρ of individuals would recover in both the treatment and control group, the actual disincentive effects from disability benefits on the labor response to mental health recovery is thus equal to $\frac{\beta_2}{\rho}$. As ρ is less than or equal to 1, the estimated effect is a lower bound. The estimated effect can thus also be interpreted as an Intention To Treat (ITT) effect.

It is important to note that the parallel trend assumption does not impose a non-anticipation assumption. Individuals are allowed to anticipate their recovery, as long as the trends would remain similar in absence of recovery, and thus in absence of anticipation of recovery. Antici-

Figure 6: Employment rates relative to recovery for the various DI subgroups

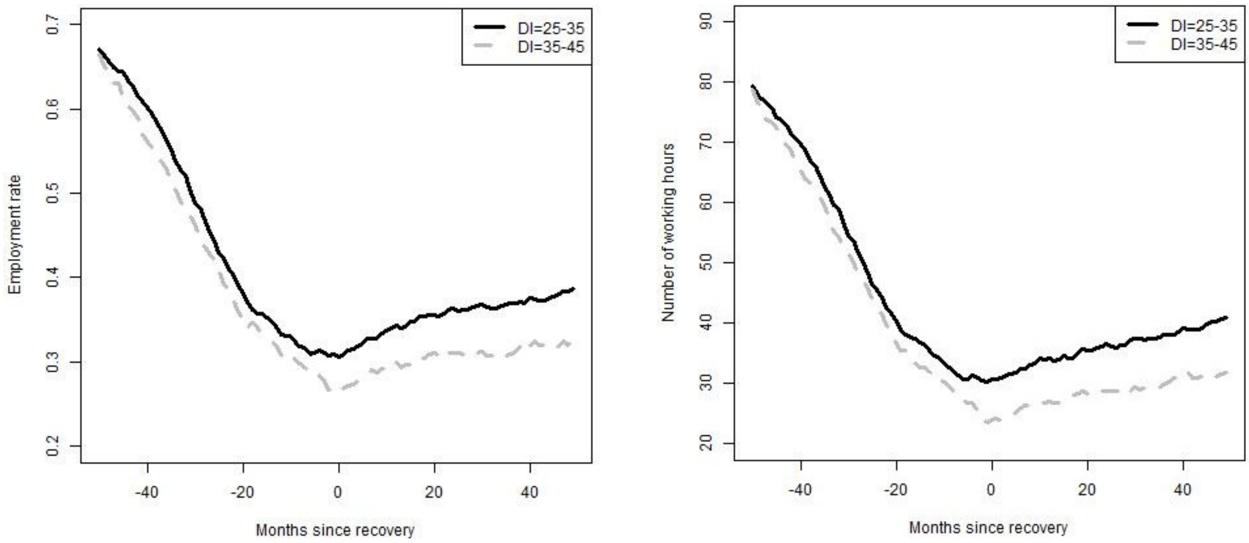


pation could cause pre-recovery divergence between the treatment and control group. To test for anticipation, various time-windows and donuts will be used. The common trend assumption imposes a certain distribution of the pre-recovery divergence to the pre-recovery difference in means and the post-recovery difference in means. If the time window is equal to the divergence period, approximately half of the divergence will be attributed to the pre-recovery difference in mean and half will be attributed to the post-recovery difference in mean, and thus to the treatment effect. As the time window grows or if a donut is incorporated, a larger part of the pre-recovery divergence will be attributed to the effect of recovery. A decreasing time window causes a smaller part of the pre-recovery divergence to be attributed to the effect of recovery.

4.1 Parallel trends

A natural starting point for selecting the control and treatment groups are the disability classes as determined in the DI application. We thus divide the sample in three group: the non-disabled, the partially disabled without DI benefits and the partially disabled with DI benefits. The non-disabled are the DI applicants whose degree of disability is assessed as zero percent. The partially disabled without DI benefits are those applicants whose degree of disability is assessed as being between zero and 35 percent, and the partially disabled with DI benefits are those applicants whose degree of disability is assessed as being between 35 and 80 percent. The non-disabled are treated as a separate group as their application outcome can be considered a corner solution: their remaining earnings capacity is assessed as at least as much as their pre-disability earnings. This group is also relatively large (9502 applicants). The partially disabled

Figure 7: Employment rate (left) and number of working hours (right) relative to recovery



without DI benefit group is comprised of 9498 individuals and the partially disabled with DI benefit group is comprised of 6078 individuals.

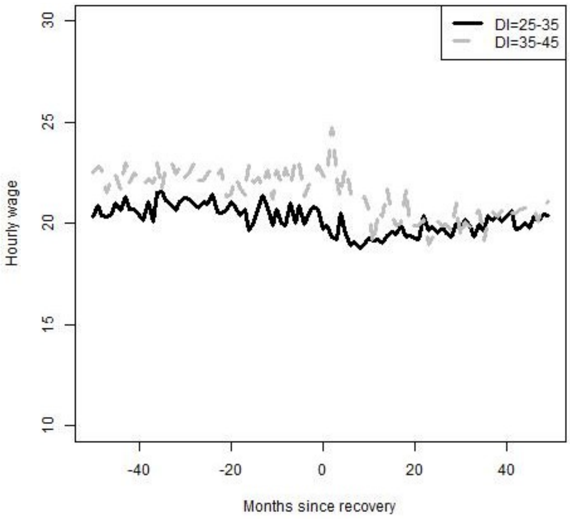
As the main assumption of the DiD estimation is the parallel trend assumption, Figure 6 shows the employment rates of the three groups relative to recovery. The general trend of all three groups is similar: a deterioration of the employment rate in the years leading up to recovery, and a stabilisation after recovery. The pre-recovery trends do however differ significantly between all groups.[5] This observation is perhaps not surprising, as the individuals with a loss of earnings capacity close to 80 percent are very different than individuals with a loss of earnings capacity close to zero percent. We therefore zoom in closer to the 35 percent disability benefit cut-off and use those with an assessed loss in earnings capacity between 25 and 35 percent as the control group, and those with an assessed loss in earnings capacity between 35 and 45 percent as treatment group. As a robustness check we also consider larger bandwidths.

Figure 7 shows the trends in the employment rate (left) and the average number of working hours (right) for the control and treatment group relative to recovery and Figure 8 shows the trends in average hourly wages. The trends and levels of both employment and average number of working hours are very similar for the control and treatment group leading up to recovery with a relatively constant and small pre-recovery difference between both groups. The trends start to diverge approximately 6 months before recovery. Given that the average duration of the mental health treatment is approximately three years, this implies that the divergence starts towards the end of the treatment. After the end of treatment, the trends diverge indicating a differential causal impact of recovery on the two groups.

The common trend for hourly wage (Figure 8) is not as strong as the common trend in employment and hours. The wage patterns follow the same general trend, but there appears to

[5]The main characteristics of the three groups also differ significantly, as can be seen in the balancing table in Appendix A.3

Figure 8: Average hourly wage relative to recovery



be more noise than in the employment and hours trend. Furthermore, the average pre-recovery wage level is higher for those with benefits. This corresponds to the fact that the pre-application hourly wage of those with benefits is higher by construction. After recovery, the wage levels converge. The convergence is largely due to a decrease in the average hourly wage of those receiving benefits. This could indicate that those who do recover, are willing to accept a lower wage given that their wage is supplemented by the DI benefits.

The visual inspection of the parallel trends seems to indicate that the parallel trend assumption holds. The parallel trends assumption can however also be tested indirectly, by testing whether the control and treatment group are comparable in terms of observable characteristics. Comparability makes it more likely that the groups would have followed similar trends in the abscence of treatment. Table 2 shows the descriptive statistics for the 25-35 and 35-45 groups, and tests for equality of means. Both groups are similar in terms of gender and nationality, but the group without benefits is on average slightly younger and better educated. The length of mental health treatment is approximately one 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, which is inherent to the assessment procedure[6].The change of health improvement, as assessed during the DI application, is very similar for both groups. Lastly, both groups have similar healthcare expenditures in the year of their DI application.

An additional, time specific, variable which should be comparable between both groups is the timing of application relative to the moment of recovery. This ensures that differences between the treatment and control group are not caused by the application procedure. Figure 9 therefore shows the survival functions for both groups. Survival is defined as not recovering, and the

[6]Given the 35% threshold used, non-manipulation would increase the comparability between the two groups. However, a significant discontinuity at the threshold is observed (Appendix A.4)

Table 2: Descriptive statistics of the control and treatment group for the DiD estimation

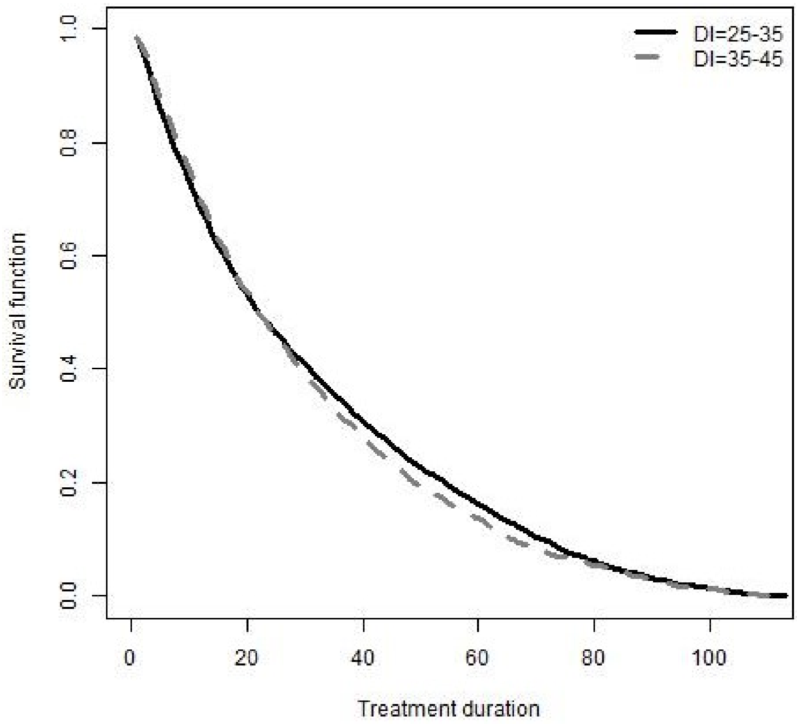
	Degree of disability:		
	25-35	35-45	P-val[a]
Age	48.460	49.401	0.001
Female	0.524	0.518	0.772
Dutch native	0.668	0.683	0.276
Education:			
Unknown	0.097	0.136	0.000
Low	0.251	0.246	0.662
Middle	0.414	0.372	0.003
High	0.238	0.246	0.518
Employment:			
Pre-application wage	16.758	17.779	0.000
Pre-application hours	33.274	33.949	0.003
Mental health:			
Treatment length[b]	32.649	34.097	0.022
DI application:			
FML	9.851	11.685	0.000
Disability percentage	29.709	40.165	0.000
Post-application wage[c]	12.442	12.319	0.018
Post-application hours[d]	32.040	29.052	0.000
Chance of health improvement[e]:			
NA	0.264	0.265	0.980
Reasonable to good	0.656	0.671	0.282
Small	0.077	0.062	0.046
Non-existent	0.003	0.003	0.724
Medical expenditures:			
Mental healthcare expenditures[f]	3028.618	3193.931	0.624
Physical healthcare expenditures[f]	1534.244	1558.932	0.8217
Observations	3270	1788	

Note: (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 the UWV, (d): potential post-application weekly working hours as determined by the UWV, (e): estimated chance of health improvement as determined by the UWV, (f) cost in euros in the year of DI application

treatment duration is the time between recovery and applications. The survival functions differ slightly, but a long rank test shows that the difference is not significant (p=0.7) (Harrington & Fleming, 1982). Furthermore, the small difference in the survival functions occurs after 20 months, and is therefore not included in the majority of the analysis.

Significant differences between the two groups thus do exist, but these differences are small in magnitude. To further test for comparability between the two groups, an DiD model will be estimated in which no control variables are included. If the treatment and control group are indeed comparable, the results of the model without control variables should be similar to those of the baseline model. Differences between the outcomes of the two models point at potential unobservable differences between the control and treatment groups.

Figure 9: Survival functions of recovery relative to application

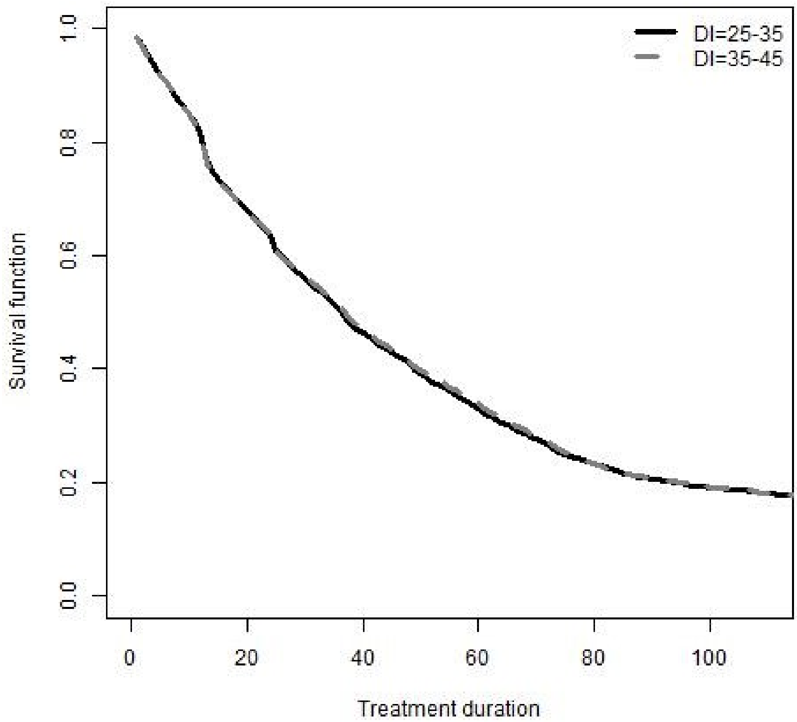


4.2 Equal occurrence of recovery

The identification of our DiD estimate is based two assumptions: the parallel trends assumption and equal occurence of recovery. Unlike in standard DiD studies, actual treatment is not observed in the current setup. Differences in the employment outcomes should however not be caused by different recovery rates between the two groups. Different recovery rates could cause divergence of the trends, even if the individual responses to recovery are similar. If the end of treatment would imply recovery for a larger proportion for a certain group, and if recovery would indeed influence employment, the group with a higher recovery rate would see a larger group average response. Testing equal occurrence of recovery is however less straightforward as recovery is not observed. Equal occurrence of recovery should hold for recovery due to treatment and for recovery due to reverse causality. The estimated chance of improvement, according to the UWV (Table 2), is very similar for both groups. Given the similar chances of improvement, differences in reverse causality would cause differences in the total recovery rates. The group affected most by reverse causality would be more likely to recover due to finding a job, causing overall recovery to be higher. This would however also imply that the exit rate from mental health treatment would be higher. To test this, Figure 10 shows the survival function, i.e. the fraction of individuals still receiving mental health treatment as a function of time, for the two groups. A long rank test shows no significant difference between the survival functions (Harrington & Fleming, 1982). As the survival functions are almost identical, it is unlikely that recovery rates due to treatment and due to reverse causality are very different.

To further investigate whether differences between the control and treatment group are

Figure 10: Survival functions of mental health treatment



driven by reverse causality, the DiD model is also estimated on subsamples with specific mental health problems: those with mood disorders, anxiety disorders or personality disorders. The estimates for anxiety disorders and personality disorders should be robust to reverse causality as a relationship between these disorders and employment has not been found (Paul & Moser, 2009). Additionally, these more homogeneous subsamples increase the comparability between the treatment and control group.

The last assumption of the DiD framework is the Stable Unit Treatment Value Assumption (SUTVA). This assumption implies that there is no relevant interaction between the members of the population. In particular, it is required that the assignment of treatment of an individual does not affect the assignment of treatment of another individual. Furthermore, treatment of an individual should not have an effect on other individuals. In the current setup this thus implies that recovery of an individual should not affect the recovery process of other individuals, and that the recovery of an individual should not affect any relevant aspects of other individuals. It is unlikely that recovery will affect the recovery of others in the sample. One could argue that, if recovery leads to increased employment, recovery of an individual increases the supply of labor. It would therefore be more difficult for others to find employment due to the recovery of this particular individual. Given that recovery happens uniformly over time, and given the national scale considered in this paper, these labor supply effects are assumed to be negligible.

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

	Employment	Hours	Wage[a]
DiD estimate	-0.015**	-2.650**	0.660*
	(0.002)	(0.240)	(0.312)
Pre-recovery difference	-0.029**	-4.590**	1.304**
	(0.002)	(0.270)	(0.301)
Recovery	0.006**	1.450**	-0.114
	(0.001)	(0.115)	(0.242)
Pre-recovery trend	-0.002**	-0.373**	0.056*
	(0.000)	(0.015)	(0.019)
Post-recovery trend	0.004**	0.424**	-0.095**
	(0.000)	(0.014)	(0.015)

Note: (a): hourly wage, , standard errors in parentheses, * significant at a 5% significance level, **significant after applying a $\frac{1}{36}$ Bonferroni correction factor

4.3 Estimation results

The estimates of the baseline DiD models are shown in Table 3.[7] The estimated causal treatment effect on employment is 1.5% points (column 1). This implies that having disability benefits reduces the effect of recovery on employment, relative to not having disability benefits. The associated effect on number of hours worked is a reduction of 2.7 hours (column 2). There is a positive wage effect (column 3) indicating that the wage of those receiving disability benefits grew more (or decreased less) due to recovery. The pre-recovery difference in means is estimated at 2.9% points and 4.6 hours respectively. The instantaneous recovery effect on the control group (0.6% points and 1.5 hours) is positive and significant.

Given the job search frictions (and potential measurement error), the recovery estimate is most likely an underestimate of the actual recovery effect. Part of the recovery effect is probably absorbed by the post-recovery time trend. To assess the magnitude of the disincentive effect of DI in comparison to the actual effect of recovery, we perform a back-of-the-envelope calculation. Assuming that the employment rate would stabilize in absence of recovery, the 12 month effect of recovery on those without disability benefits would be equal to the recovery coefficient plus the 12 months trend effect. This thus equals 5.4% points[8] in terms of employment (and similarly 6.5 working hours[9]). The effect of recovery on those with disability benefits equals 3.9% points[10] in terms of employment (and similarly 3.8 working hours[11]).

Given the significant results found in the baseline models, different specifications were used to assess the robustness. Mean level estimation, non-parametric estimation, exclusion of control variables and using monthly dummies yield very similar results compared to the baseline model (Appendix A.5). The standard errors of the baseline model, mean level estimation model and

[7]We use a $\frac{1}{36}$ Bonferroni correction factor as we test a total of 36 DiD estimates. Assuming independence, and given that all estimates are significant, the actual significance level is higher than 5% The probability that all estimates for a certain measure are significant, while actually being zero, is $0.05^{36} = 0.1 * 10^{-48}$.

[8]Instantaneous recovery effect plus 12 times the post-recovery time trend $= 0.6 + 12 * 0.4 = 5.4$

[9]Instantaneous recovery effect plus 12 times the post-recovery time trend $= 1.450 + 12 * 0.424 = 6.5$

[10]Effect of those without benefits minus the disincentive effects of DI benefits $= 5.4 - 1.5 = 6.5$

[11]Effect of those without benefits minus the disincentive effects of DI benefits $= 6.5 - 2.7 = 3.8$

Table 4: Robustness specifications DiD for employment, hours and hourly wage

	Specification				Outcome measures		
	Window[a]	Donut	N_c^b	N_t^c	Employment	Hours	Wage[d]
Baseline model	12	0	3270	1788	-0.015**	-2.650**	0.660*
					(0.002)	(0.207)	(0.312)
6 month window	6	0	3270	1788	-0.008**	-1.696**	1.345*
					(0.002)	(0.236)	(0.552)
48 month window	48	0	3270	1788	-0.022**	-4.041**	0.642**
					(0.001)	(0.141)	(0.072)
Donut	48	12	3270	1788	-0.025**	-4.548**	-1.052**
					(0.002)	(0.193)	(0.082)
20%-50%	12	0	4848	2484	-0.026**	-3.725**	0.338
					(0.002)	(0.173)	(0.241)

Note: (a) Incorporated number of months before and after recovery , (b): number of individuals in the control group, (c): number of individuals in the treatment group, (d): hourly wage, standard errors in parentheses, * significant at a 5% significance level, **significant after applying a $\frac{1}{36}$ Bonferroni correction factor

non-parametric estimation model are similar implying that clustered standard errors deal with potential serial correlation sufficiently well. Given that the estimates are not affected by the exclusion of the control variables, it is likely that the control group and treatment group are indeed very similar in all relevant aspects. Using monthly dummies instead of a linear time trend does not affect the estimates either, implying that the linear time trend is adequate, or that deviations from the linear trend are similar for the treatment and control group. Table 4 shows the estimated treatment effects of the various robustness specifications. For comparison, the first row shows the results of the baseline models.

The estimates for employment and number of working hours are robust to the various specifications used. Shortening the time frame considered decreases the DiD estimates (row 2). As mentioned, a shorter time frame implies a large part of the pre-recovery divergence is attributed to the pre-recovery difference in means, and a smaller part is attributed to the treatment effect. As expected, expanding the time frame to 4 years (48 months), and by doing so focussing on the long term, increases the effect on employment and hours to 2.2% points and 4 hours respectively (row 3). Including a one year symmetric donut, and thereby assuming that all divergence around the recovery date is fully attributable to recovery, results in an upperbound of the DiD estimates (row 4). The estimated effect on employment is 2.5% points and the effect on number of working hours is 4.5 hours. Relative to an employment rate of approximately 30% points, and an average number of working hours of 30 hours, it implies an effect of approximately 10% and 15%. Given the larger relative effect on the number of working hours, there is most likely an intensive and extensive effect. Having disability benefits reduces the number of individuals that start working due to recovery, and reduces the number of individuals who increase their number of working hours due to recovery. The estimates are also robust to including individuals further away from the 35% disability threshold (row 5).

The results for hourly wages not robust to the various specifications. Not only the mag-

Table 5: DiD estimates for employment, hours and hourly wage on various subsamples

	Specification				Outcome measures		
	Window[a]	Donut	N_c^b	N_t^c	Employment	Hours	Wage[d]
Baseline model	12	0	3270	1788	-0.015**	-2.650**	0.660*
					(0.002)	(0.207)	(0.312)
Anxiety disorders	12	0	410	167	-0.026**	-5.589**	0.118
					(0.007)	(0.748)	(0.447)
Personality disorders	12	0	540	282	0.014*	-1.194*	-0.721
					(0.005)	(0.556)	(0.438)
Mood disorders	12	0	1294	744	-0.034**	- 5.030**	-0.041
					(0.003)	(0.306)	(0.564)

Note: (a) Incorporated number of months before and after recovery , (b): number of individuals in the control group, (c): number of individuals in the treatment group, (d): hourly wage, standard errors in parentheses, * significant at a 5% significance level, **significant after applying a $\frac{1}{36}$ Bonferroni correction factor

nitude, but also the sign (and significance) of the estimate depends on the specification used. The baseline, 6 months and 48 months specifications result in a positive effect on the wage, whereas the use of a donut results in a negative estimate. Including more individuals renders the estimate insignificant (despite stronger power). Given the results of the various robustness specifications and the noisy trends as shown in Figure 8 , the DiD estimate for wage should be interpreted with caution.

To rule out that the results are driven by differences in reverse causality, Table 5 shows the results of the baseline models estimated on smaller subsamples. There is large heterogeneity in the DiD estimates based on the mental health diagnoses. The estimates for the largest group, those diagnosed with mood disorders, are larger than the baseline estimates. As mentioned, reverse causality could be present for this subsample. For the other two subsamples reverse causality is less likely. The estimated effects on employment and number of hours worked for those with anxiety disorders are large and significant as well. Surprisingly, the results for those with personality disorders are positive and borderline significant. This implies that those with personality disorders are more likely to return to the workforce upon recovery when receiving disability benefits, relative to their counterparts without disability benefits. Given the large effects for the subsamples with anxiety and mood disorders, it is unlikely that all effects are driven by reverse causality.

Summing up, there are significant disincentive effects of having disability benefits on the employment response to mental health recovery. The estimated effects range between 1.5 to 3.4% points in terms of employment, and between 2 and 6 working hours. Part of these effects accumulate before the end of treatment (given the effect of incorporating a donut), implying that individuals either anticipate their recovery, or that recovery happens before the end of treatment. As the relative effects on the average number of working hours is larger than the relative effect on the employment rate, there is both an intensive and extensive effect: more individuals in the control group start working, and those who already work increase their number of working hours to a greater extent, compared to the treatment group. Given the robustness of the results, it seems unlikely that the results are fully caused by reverse causality. To assess the strength of the proxy used for recovery and to determine whether the results extend to non-mental health, the following subsection performs the same analysis using drops in mental healthcare expenditures and drops in non-mental healthcare expenditures as proxies for recovery.

4.4 Analysis based on healthcare expenditure

The analysis so far used the end of mental health treatment as a proxy for (mental) health improvement. As a robustness analysis, we can alternatively use a drop in healthcare expenditures to proxy recovery. To be precise, the year of recovery (the healthcare expenditure data is only available on a yearly basis) is defined as the first year in which expenditure drops below, and stays below, 20%[12] of the healthcare expenditures in the year before the DI application. The year before the DI application is used as the baseline, because this year constitutes the DI waiting period in which health status is generally at its lowest. Healthcare expenditures are split into mental healthcare expenditures and non-mental healthcare expenditures (see Appendix A.2) which results in two different proxies: a proxy for mental health recovery and a proxy for physical health recovery. Given that we only have yearly healthcare expenditure data, recovery could either be in (the beginning of) the first year with low expenditure, or in the year prior to the first low-cost year.

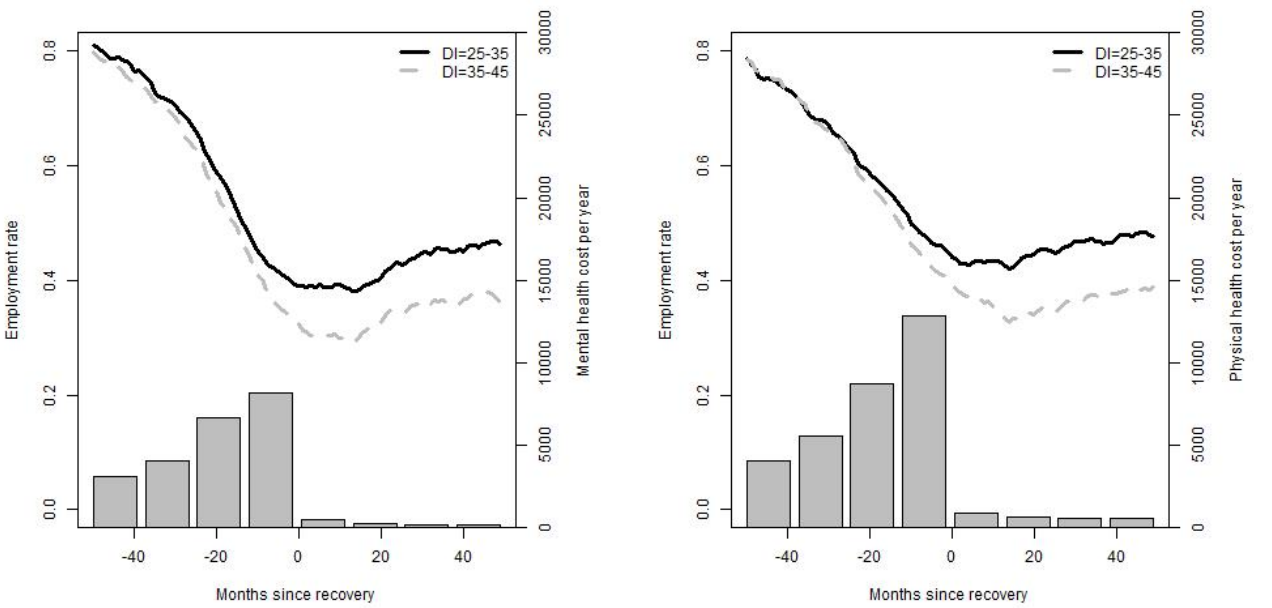
The sample selection is similar to the sample selection used above. We select individuals for who we observe a drop in expenditures. Furthermore, the DI application should be due to mental health problems (in the analysis based on mental health) or due to non-mental health problems (in the analysis based on physical health) to ensure that recovery aligns with the reason of disability. The resulting sample for mental health is comprised of 7773 individuals and the sample for non-mental health is comprised of 10162 individuals.

Figure 11 shows the employment rates relative to recovery for the two different proxies (left axis). The bars indicate the average mental or physical healthcare expenditures for each year (right axis). Prior to recovery, expenditures increase to approximately 8000 and 13000 euros per year. After the drop, the cost remain low (and decrease further). The trends in the employment rates are very similar for the control and treatment group. Divergence starts approximately 1 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 shows the DiD estimates for both proxies. Panel A shows the results for recovery based on a drop in mental health and Panel B shows the results for recovery based on a drop in physical health. The baseline specifications, shown in the first rows of both panels, incorporate 24 months before and 24 months after recovery without a donut. However, given that recovery could occur in the first low-cost year, or in the year prior to the first low-cost year, the second rows in both panels incorporate a donut of 12 months before and 12 months after recovery,

[12]Results for other thresholds are similar(see Appendix A.5)

Figure 11: Employment rates relative to recovery for mental health (left) and physical health (right)



thereby excluding the two years in which recovery could potentially occur. The estimates for mental health are larger than those obtained through the other proxy. This could be due to the fact that the mental health treatment used above is a subset of the mental health treatment contained in the cost data. The aligns with the observation that the post-recovery time trend is also greater for the proxy based on healthcare expenditures.

The estimates for physical health (Panel B) are larger in magnitude than those for mental health (Panel A). This could be due to the fact that a drop in healthcare expenditures is more likely to correspond to health improvement for physical health problems, or it could be due to the fact that the labor response to recovery, and therefore the disincentive effect of DI benefits on this response, are larger for those with a physical health improvement.

5 Economic model

The focus of this paper is on the disincentive effects of disability benefits on the labor response to mental health recovery. However, the estimates obtained in the DiD analysis are ITT estimates. Transforming these ITT estimates into actual recovery effects is not possible, as it is not known what fraction of the sample actually recovers. To gain insights into the maximal effect of actual recovery, a different approach is therefore required. A structural labor supply model is estimated, in which recovery is defined as regaining ones pre-disability earnings capacity. Based on the loss in earnings capacity due to disability, as assessed by the UWV, labor supply effects of recovery can be estimated. As the disability benefits continue after recovery for disability beneficiaries, the model can be used to assess the disincentive effects of disability benefits.

Table 6: DiD estimates based on healthcare expenditures for employment, hours and wage

	Specification:						
	Window^a	Donut	N_c^b	N_t^c	Employment	Hours	Wage^d
Panel A: Mental health							
Baseline specification	24	0	4879	2894	-0.044**	-6.143**	-1.127**
					(0.001)	(0.158)	(0.097)
Donut specification	24	12	4879	2894	-0.057**	-8.492**	-1.631**
					(0.003)	(0.307)	(0.173)
Panel B: Physical health							
Baseline specification	24	0	6212	3950	-0.052**	-8.338**	-0.307**
					(0.001)	(0.135)	(0.064)
Donut specification	24	12	6212	3950	-0.079**	-12.517**	-0.626**
					(0.002)	(0.260)	(0.114)

Note: (a) Incorporated number of months before and after recovery , (b): number of individuals in the control group, (c): number of individuals in the treatment group, (d): hourly wage, standard errors in parentheses, * significant at a 5% significance level, **significant after applying a $\frac{1}{36}$ Bonferroni correction factor

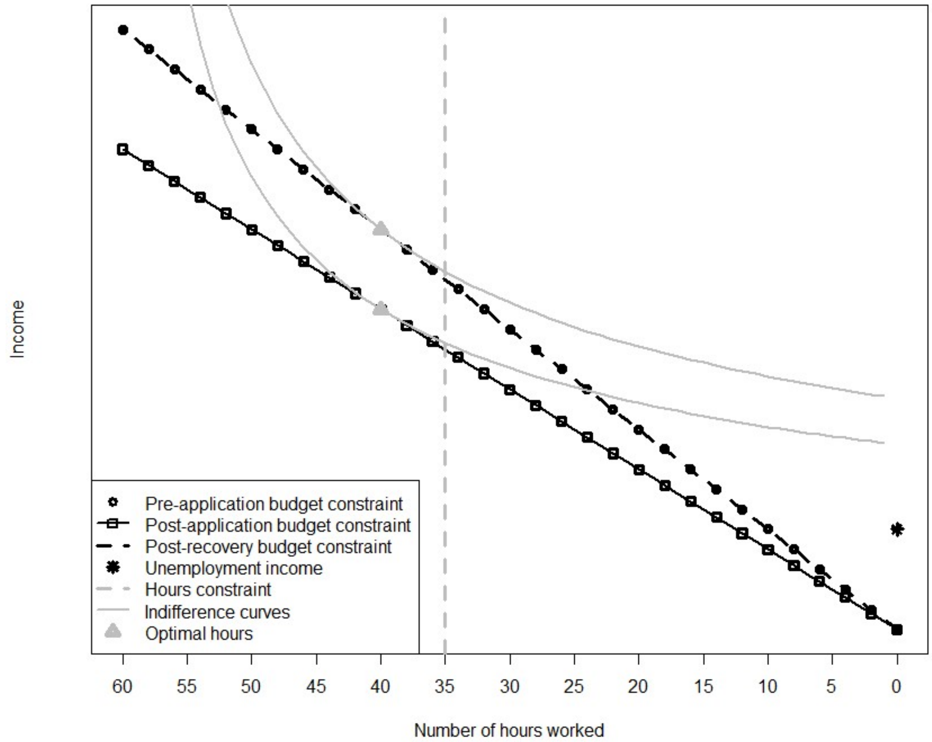
The structural labor supply model is consists of three stages. In the first stage, we estimate utility preferences for work and leisure for each individual in the sample based on the hours decision before disability. In the second stage, the onset of the disability implies a loss in earnings capacity along two observed dimension: the maximum number of working hours and the hourly wage. Together with the receipt of DI benefits for those who have been awarded benefits, this loss in earnings capacity changes the budget line, possibly resulting in a different optimal number of working hours. The third stage models recovery, through restoring the earnings capacity to the pre-disability level. For those who have not been granted disability benefits, the third stage is thus identical to the first stage. For those who who have been granted disability benefits, the budget line is different, resembling the disincentive effects of disability benefits.

The model does not take any labor demand aspects into consideration and tax effects are not taken into account when constructing the budget constraints. Hourly wages are assumed to be exogenous, and given hourly wages, individuals choose their working hours. Employment decisions therefore follow from utility maximization. The model assumes that the utility function over labor and income stays constant over time. It is thus assumed that both being awarded benefits and changes in health do not affect the utility function. The following subsections will show a graphical illustration of the model and the setup of the model, after which the results will be discussed.

5.1 Graphical illustration

This graphical illustration considers two hypothetical individuals. Both individuals are assumed to work full-time (40 hours) before their disability benefit application. Continuation benefits are 28% of the statutory minimum wage. For simplicity, unemployment benefits are set equal to

Figure 12: Budget constraints of an individual who has not been awarded disability benefits



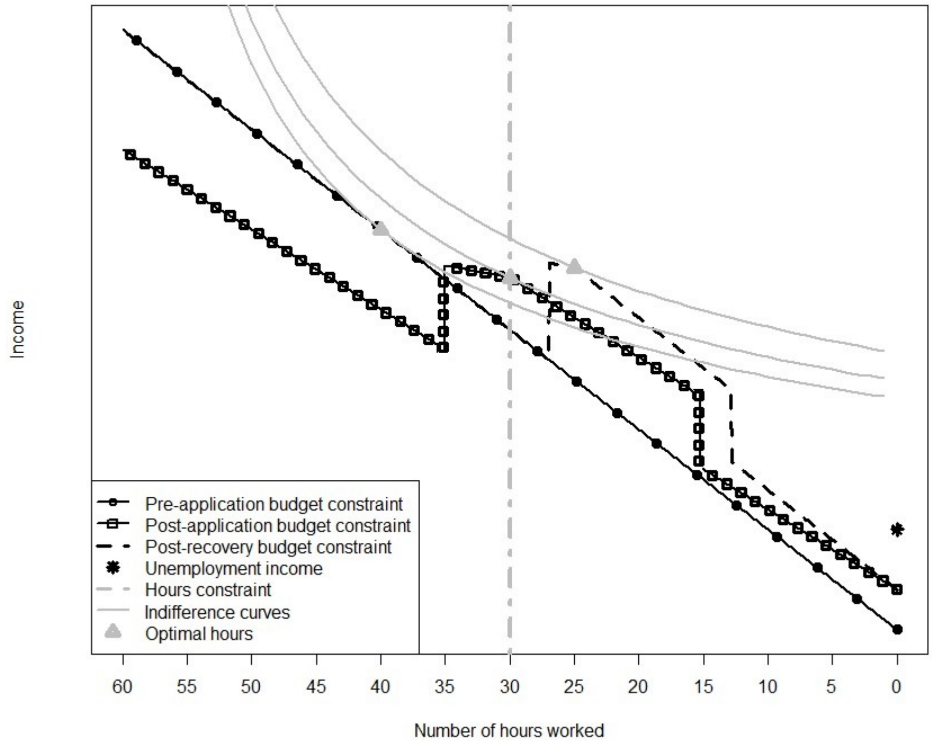
poverty relief, which is 70% of the minimum wage. As income is supplemented up to the point of poverty relief, the poverty relief income level is included in all budget constraints. The first individual has an assessed degree of disability of 30%, and will thus not be awarded disability benefits. The degree of disability is comprised of an hours restriction of 87.5% and an hourly wage restriction of 80%. The second individual has an assessed degree of disability of 40% and will thus be awarded disability benefits. The degree of disability is comprised of an hours restriction of 75% and an hourly wage restriction of 80%.

Figure 12 shows the budget constraints for the first individual who has not been awarded disability benefits. The pre-application budget constraint, shown by the line with dots, is a straight line in which income increases one-to-one with wage. The only non-linearity in the pre-application budget constraint is caused by the unemployment income. 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[13]) 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 DI assessment.

Figure 13 shows the budget constraint for the second individual in the pre-application, post-

[13]The gray indifference curves are based on a Cobb-Douglas utility function.

Figure 13: Budget constraints of an individual who has been awarded disability benefits



application and post-recovery phase. The pre-application budget constraint, shown by the dots, is identical to the pre-application budget constraint of the first individual. It increases one-to-one with wage and has one discontinuity due to unemployment income. 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%. Furthermore, the disability benefits cause non-linearities at 50% of the remaining earnings capacity, 100% of the remaining earnings capacity and 65% of the pre-application income[14]. In the post-application phase, the hours constraint, as shown by the gray dashed line, is enforced. Lastly, the post-recovery budget constraint is shown by the black dashed line. Recovery implies that the slope of the budget constraint is identical to the slope of the pre-application budget constraint. However, the disability benefits cause similar discontinuities in the budget constraint as in the post-application budget constraint. Given the higher hourly wage, these discontinuities occur at a lower number of hours worked. As can be seen from the optimal number of hours, the response to recovery is a further reduction in the number of working hours.

For both individuals, a utility parameter can be estimated by finding the utility function tangent to the budget constraint at the observed number of working hours in the pre-application phase. Given this utility function, the optimal post-application and post-recovery employment level can be computed through optimizing over the post-application and post-recovery budget

[14]The disability benefits are linked to the statutory minimum income until half of the earnings capacity is used. If 50 to 100 % of the earnings capacity is used, benefits are linked to the pre-application wage. Once the earnings capacity is exceeded, benefits are reduced, and fully terminated once earnings equal 65% of pre-application earnings. See section 3 for a detailed description.

constraints respectively. The response to recovery can then be computed as the difference between the optimal post-application and the optimal post-recovery employment level.

Similar changes in hourly wages can lead to different responses, based on whether or not an individual has been awarded disability benefits. The model will exploit these differences to estimate the hypothetical disincentive effects of disability benefits on the response to recovery.

5.2 Model setup

As mentioned, the model assumes utility maximization over the number of hours worked. The utility function used is a Cobb-Douglas specification in which the utility weights are normalized to one[15]. A base level income will be used for those who do not work, to incorporate the unemployment insurance system in the Netherlands. The model uses three different points in time: pre-application, post-application and post-recovery. The general utility maximization problem is as follows:

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

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

$$I_i(E_i) = \max(E_i w_i + DI(E_i), UI) \tag{5}$$

In which L_i is the amount of time available for leisure. Total available time, T, can be allocated to either leisure or work, E_i. Furthermore, total income, I_i, equals income from labor plus potentially income from disability benefits. Labor income equals the amount of working hours, multiplied by the hourly wage, w_i. The amount of disability benefits, DI, potentially depend on the number of working hours through earnings. Both individuals with and individuals without disability benefits can apply for social assistance, and total income will always be supplemented to the level of social assistance. Social assistance payments are denoted by UI, as the level is equal to the long-term Unemployment Income. After substituting the constraints and taking the natural logarithm, maximization with respect to the number of hours gives the following First Order Condition (F.O.C.):

$$\frac{\lambda_i}{T - E_i} = \frac{(1-\lambda_i)(w + DI'(E_i))}{E_i w_i + DI(E_i)} \tag{6}$$

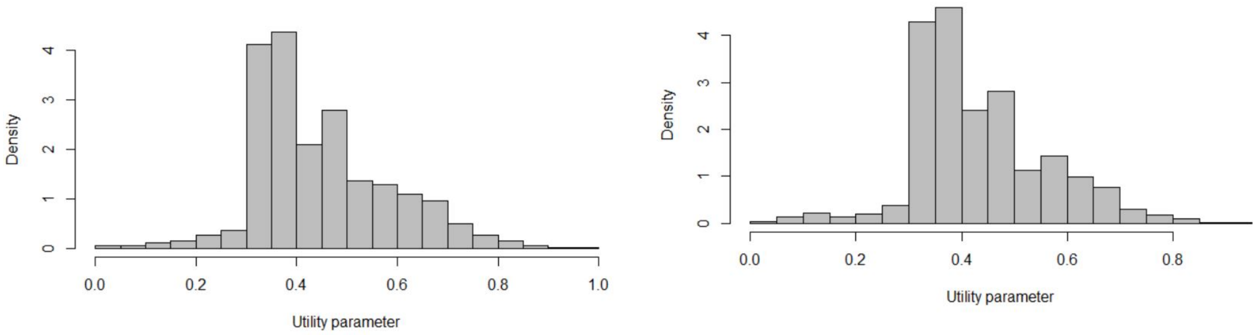
In which $DI'(E_i)$ is the derivative of the disability benefits with respect to number of hours. Given hourly wages and number of working hours, the preference parameter can be estimated as

$$\lambda_i = \frac{(T - E_i)(w_i + DI'(E_i))}{DI(E_i) + Tw_i + DI'(E_i)(T - E_i)} \tag{7}$$

Given the preference parameter λ_i and the wage, the optimal number of hours can be

[15] Given that the utility function is determined based on one employment decision, at most one preference parameter can be solved. A Cobb-Douglas utility function with weights normalized to one is therefore used.

Figure 14: Distribution of the pre-application utility parameter for those without benefits (left) and those with benefits (right)



determined. There is however no closed form solution for the optimal E_i^*. In case disability benefits are awarded, the budget constraint is piece-wise linear. The optimal number of hours will be determined for every section of the budget constraint, and the resulting utilities are compared to obtain the optimal number of working hours. Finally, the utility from working the optimal number of working hours is compared to the utility of not working, to determine the actual employment decision of an individual. The resulting optimization problems at the various stages of the model and their solutions are shown in Appendix A.6.

5.3 Results

The sample used for the estimation of the structural labor supply model is almost identical to the sample used for the DiD estimation. 48 individuals in the DiD sample are excluded, as either the hours restriction or the hourly wage restriction is not given. Given that the observed pre-application number of working hours ranges between 0 and 60, the number of hours to be divided between leisure and work, T, has been set at 60. Furthermore, the unemployment income has been set at 70% of the 2019 statutory minimum income, i.e. $0.7 * €377 = €264$.

Based on the information on pre-application hourly wages and working hours, we infer a distribution of the preference parameter λ as shown in Figure 14. The bunching around 0.4 resembles individuals working approximately 40 hours. The mean of the distribution for the group with benefits is slightly lower than the mean of the distribution of the group without benefits due to the fact that those with DI benefits work more hours before their application.

To estimate the post-application optimal number of working hours, based on the pre-application utility parameter, the post-application hourly wage is required. The hourly wage predicted by the UWV will be used as a proxy for the actual post-application hourly wage. In case the predicted hourly wage is above the pre-application hourly wage, the post-application hourly wage is set equal to the pre-application hourly wage. This is done to ensure that recovery does not imply an hourly wage decrease. To assess the fit of the model, the prediction error in the post-application number of working hours is calculated. The predicted number of working hours is on average 6 hours more than the actual number of hours, which is mainly caused by an

Table 7: Estimated response to recovery and DI benefit effect

25-35		35-45				DI benefit effect	
		With benefits		Without benefits			
Employment	Hours	Employment	Hours	Employment	Hours	Employment	Hours
23.2%	8.2	5.0%	-2.4	21.8%	7.7	16.8%	10.1

over-prediction of the number of individuals who would work full time. The hours restriction is binding for 23% of those without disability benefits, and for 19% of those with disability benefits.

Lastly, the response to recovery can be estimated (Table 7). To ensure that differences in the response are not caused by differences in the post-application fit, the response to recovery is defined as the predicted post-recovery number of working hours, minus the predicted post-application number of working hours. According to the model, 23.2% points of those without benefits start work because of recovery, compared to 5.0% points of those with benefits. The effect of recovery on the average number of working hours is an increase of 8.2 hours for those without benefits, and a decrease of 2.4 hours for those with benefits. The difference in the response to recovery between the two groups is thus 18.2% points in terms of employment and 10.6 working hours.

To determine what part of this difference is attributable to having disability benefits, and what part is due to inherent differences between the groups, the recovery response of the group with disability benefits, in case they had not been awarded disability benefits, is computed. The recovery response for this counterfactual case can be determined by solving the pre-application optimization problem, using the post-application hourly wages. The estimated recovery response is an increase in the employment rate of 21.8% points, and an increase of 7.7 working hours. If those with benefits, would not have been awarded benefits, their response to recovery would be similar to the response of those without benefits. Given these similar responses, the group without benefits is a suitable control group in the DiD estimation.

The disincentive effects of disability benefits on the labor response of those with disability benefits, defined as the difference between their response in case they would have, or would not have benefits, is 16.8% points in terms of employment and 10.1 working hours. These estimates should be seen as upper bound, as the model assumes full recovery and does not take any labor demand aspects into account.

The structural labor supply model illustrates that having even partial disability benefits, greatly affects the response to recovery. On a monthly basis (as used in the DiD estimation), the effect of disability benefits is a reduction in the employment rate of 16.8% points and a reduction in the monthly number of working hours of approximately 44.4 hours (10.1 x 4.4). These estimates are approximately ten times as large as the upper bounds of the DiD model. Given that it is likely that only a part of the DiD sample recovers, and that the presence of comorbidities imply that recovery can at most be partial, the DiD estimates are still relatively large.

6 Conclusion

This paper revolves around the question whether labor supply responses to improvements in mental health are partly eliminated by the disincentives of disability benefits. As such, the aim is to deepen the understanding on low work resumption rates of DI benefit recipients, particularly in schemes intended for the temporarily disabled. We consider employment responses to an improvement of (mental) health for DI applicants in the Netherlands with and without benefits.

Applying a difference-in-differences (DiD) framework, we estimate the impact of recovery for DI recipients, using those just below the disability benefit threshold as a control group for those just above the threshold. Conservative estimates of the disincentive effects of disability benefits indicate a reduction in the probability to be employed of 1.5 % points and a reduction in the number of working hours of 2.7 hours. Mental health recovery is proxied for by the end of mental health treatment, but an alternative definition based on significant drops in healthcare expenditures confirm these results. Furthermore, we find similar disincentives for physical (as opposed to mental) health recovery.

Given that our proxy represents partial recovery, the obtained DiD estimates are intention-to-treat estimates and thus a lower bound of the effect in case of full recovery. To shed more light on these potential effects, we consider a structural labor supply model with distinct budget constraints for individuals with and without DI benefits. The predicted theoretical disincentive effects of disability benefits are large: disability benefits reduce the response to recovery by 16.8 % points in terms of employment.

Given that there are disincentive effects of having disability benefits on the employment response to mental health improvement, a pertaining question is how the results relate to other evidence. Compared to earlier findings on disincentive effects of DI benefits, ranging from 20 to 30 % points in terms of employment (Bound, 1989; French & Song, 2014; Chen & Van der Klaauw, 2008), the obtained estimates are small. One should however take into account that in our setting, effects are expected to be smaller for two reasons. First, we estimate intention-to-treat effects as our proxy only represents partial recovery. Our model prediction for the disincentives in case of full recovery are 16.8% points, which lies closer to earlier findings. Second, the response to health recovery may be limited in any case. Earlier studies estimated the effect of (negative) health shocks to range from 7 to 15 % points in terms of employment (García Gómez et al., 2013; Lindeboom et al., 2016). These are only slightly larger than our back-of-the-envelope calculation for the employment response to mental health recovery for those without benefits (5.4% points, see section 4.3). Given this relatively small response, the disincentive for those receiving benefits is sizeable as it offsets approximately 25% of the response to recovery.

We conclude that there are disincentive effects of receiving disability benefits on the response to (mental) health improvement. These findings add to earlier studies that showed existence of general disincentive effects of having disability benefits: disability benefits not only lower the probability of being employed, they also inhibit return to the workforce upon recovery.

References

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. *NBER working paper*.

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.* (data retrieved from CBS Statline, `https://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=37789KSZ&D1=a&D2=234-274&HDR=T&STB=G1&VW=T`)

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.

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

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. *Tinbergen Institute Discussion Paper 2018-023/V*.

Godard, M., Koning, P., & Lindeboom, M. (2019). Targeting disability insurance applications with screening. *IZA Discussion Paper*.

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

Harrington, D. P., & Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika*, *69*(3), 553–566.

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. *Economisch Statistsche 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?* The World Bank.

Maestas, N., Mullen, K. J., & Strand, A. (2013). Does disability insurance receipt discourage work? using examiner assignment to estimate causal effects of ssdi receipt. *American economic review*, *103*(5), 1797–1829.

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.* (data retrieved from Social Expenditure Database, `https://stats.oecd.org/Index.aspx?DataSetCode=SOCX_DET`)

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. *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.

7 Appendix

A.1 Pre-application process

Employed

If a long-term 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, 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 public social benefit administration.
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, employee and employer make 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 wage to the employee during illness. If the employer has not put in sufficient effort to reintegrate the employee, the period in which the employer is obliged to pay 70% of the last paid wage is extended by one year. If the employee cannot return to his/her old job, the employer is required to offer suitable a suitable job. This job can also be at a different employer. If the employee refuses a suitable job, payments by the employer can be stopped, and the employee can be fired. If the employee and employer disagree about the reintegration, an expert opinion can be requested at the UWV.

Unemployed

Disability insurance in the Netherlands is an insurance against loss of income due to disability. There is however a group of individuals eligible for disability insurance without an employer. This group for example contains individuals whose (temporary) contract has ended and who receive unemployment benefits and on-call workers. For these workers, 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 the 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. The 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. The UWV will do all the tasks of the employer.

A.2 Mental health care expenditures and non-mental health care expenditures

Table 8: Construction of mental healthcare expenditures and physical healthcare expenditures

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.3 Descriptive statistics of DI groups

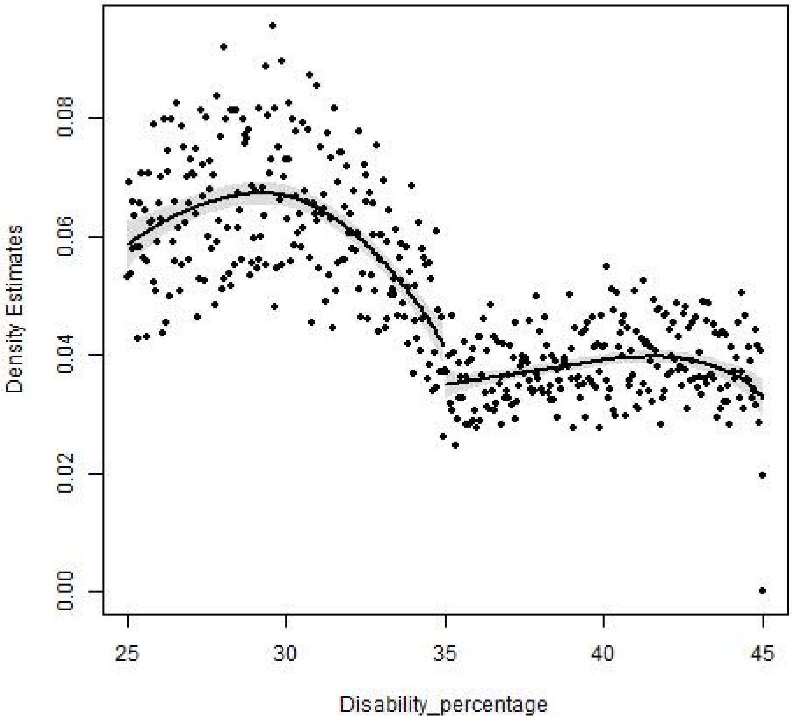
Table 9: Descriptive statistics by degree of disability

	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 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 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]	6022.112	2861.132	3542.619
Physical healthcare expenditures[f]	2194.982	1520.925	1685.192
Observations	9502	9498	6080

Note: (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 the UWV, (d): potential post-application weekly working hours as determined by the UWV, (e): estimated chance of health improvement as determined by the UWV, (f) cost in euros in the year of DI application

A.4 McCrary density test on disability threshold

Figure 15: McCrary (2008) density test on the disability threshold



A.5 Robustness specifications

Table 10: Additional robustness specifications DiD for employment, hours and hourly wage

	Specification				Outcome measures		
	Window[a]	Donut	N_c^b	N_t^c	Employment	Hours	Wage[d]
Baseline model	12	0	3270	1788	-0.015**	-2.650**	0.660*
					(0.002)	(0.207)	(0.312)
No covariates	12	0	3270	1788	-0.015**	-2.543**	0.474
					(0.002)	(0.207)	(0.311)
Monthly dummies	12	0	3270	1788	-0.015**	-2.650**	0.656*
					(0.002)	(0.207)	(0.312)
Mean level	12	0	3270	1788	-0.015**	-2.543**	0.496
					(0.002)	(0.315)	(0.405)
Non-parametric	12	0	3270	1788	-0.015**	-2.543**	0.477
					(0.002)	(0.207)	(0.311)

Note: (a) Incorporated number of months before and after recovery , (b): number of individuals in the control group, (c): number of individuals in the treatment group, (d): hourly wage, standard errors in parentheses, * significant at a 5% significance level, **significant after applying the $\frac{1}{36}$ Bonferroni correction factor

Table 11: DiD estimates based on various drops in mental healthcare expenditures

	Specification				Outcome measures
	Window^a	Donut	N_c^b	N_t^c	Employment
90 % drop	12	0	4677	2747	-0.040** (0.001)
80 % drop	12	0	4879	2894	-0.044** (0.001)
60 % drop	12	0	5273	3133	-0.047** (0.001)
40 % drop	12	0	5491	3367	-0.046** (0.001)

Note: (a) Incorporated number of months before and after recovery , (b): number of individuals in the control group, (c): number of individuals in the treatment group, standard errors in parentheses, * significant at a 5% significance level, **significant after applying the $\frac{1}{36}$ Bonferroni correction factor

A.6 Economic model

The general utility maximization problem is as follows:

$$\max_{E_i} u(L_i, I_i) = L_i^{\lambda_i} I_i^{1-\lambda_i} \tag{8}$$

$$\text{s.t. } L_i = T - E_i \tag{9}$$

$$I_i(E_i) = \max(E_i w_i + DI(E_i), UI) \tag{10}$$

In which L_i is the amount of time available for leisure. Total available time, T, can be allocated to either leisure or work, E_i. Furthermore, total income, I_i, equals income from labor plus potentially income from disability benefits. Labor income equals the amount of working hours, multiplied by the hourly wage, w_i. The amount of disability benefits, DI, potentially depend on the number of working hours. Both individuals with and individuals without disability benefits can apply for social assistance, and total income will always be supplemented to the level of social assistance. Social assistance payments are denoted by UI, as the level is equal to the long-term Unemployment Income. After substituting the constraints and taking the natural logarithm, maximization with respect to the number of hours gives the following First Order Condition (F.O.C.):

$$\frac{\lambda_i}{T - E_i} = \frac{(1-\lambda_i)(w + DI'(E_i))}{E_i w_i + DI(E_i)} \tag{11}$$

In which $DI'(E_i)$ is the derivative of the disability benefits with respect to number of hours. Given hourly wages and number of working hours, the preference parameter equals:

$$\lambda_i = \frac{(T - E_i)(w_i + DI'(E_i))}{DI(E_i) + Tw_i + DI'(E_i)(T - E_i)} \tag{12}$$

Given the preference parameter λ_i and the hourly wage, the optimal number of hours can be determined. There is however no closed form solution for the optimal E_i^*. In case disability benefits are awarded, the budget constraint is piece-wise linear. The optimal number of hours will be determined for every section of the budget constraint, and the resulting utilities are compared to obtain the optimal number of working hours. Finally, the utility from working the optimal number of working hours is compared to the utility of not working, to determine the actual employment decision of an individual.

A.4.1 Pre-application

Before the application, disability benefits have not been awarded, i.e. $DI(E_i) = 0 = DI'(E_i)$. Therefore, the utility parameter equals:

$$\lambda_i = \frac{T - E_i}{T} \tag{13}$$

This implies that the utility parameter is equal to the proportion of total time spend on leisure. The pre-application budget constraint is linear, with the only exception at $E_i = 0$. The optimal employment, given that someone works, is equal to $E_i^* = (1 - \lambda_i)T$. As mentioned, someone will work if utility from doing so exceeds the utility of not working:

$$u(E_i^*) > u(E_i = 0) \tag{14}$$

$$(T - E_i^*)^{\frac{T-E_i^*}{T}} (E_i^* w_i)^{\frac{E_i^*}{T}} > T^{\frac{T-E_i^*}{T}} UI^{\frac{E_i^*}{T}} \tag{15}$$

Given that the utility from working, the left side, is increasing in the hourly wage, whereas the utility of not working is constant, there exist some wage level at which an individual is indifferent between working and not working. For hourly wages higher than this wage level an individual will decide to work, whereas for lower hourly wages an individual will not work. In the pre-application case, this wage level is equal to:

$$\tilde{w}_i = \frac{T^{\frac{T-E_i^*}{E_i^*}} UI (T - E_i^*)^{\frac{E_i^*-T}{E_i^*}}}{E_i^*} \tag{16}$$

Hence, the decision whether or not to work depends on the hourly wage, whereas the optimal number of working hours, given that one decides to work, does not depend on the hourly wage. The pre-application information will be used to determine the utility parameters, and whether or not it is optimal for someone to work or not.

A.4.2 Post-application

After the disability assessment, the optimization problem changes. First, depending on the outcome of the assessment procedure, the DI function and thus the budget constraint changes. As illustrated, the budget constraint of individuals who have been awarded benefits becomes piece-wise linear. Furthermore, the slope of the budget constraint changes as the hourly wage an individual can earn decreases. Lastly, the maximum number of hours an individual can work decreases, creating an additional constraint.

For those who have not been awarded disability benefits, Equation (14) still holds. The optimal level of employment, given that someone works, remains at $E_i^* = (1 - \lambda_i)T$, unless this level exceeds the hours restriction, in which case the optimal level equals the hours restriction. Changes in hourly wage thus only affect the decision to work, and not the decision on the number of hours.

For those who have been awarded benefits, the location on the budget constraint depends on current earnings relative to either the remaining earnings capacity or the pre-application earnings. Denote by REC the remaining earnings capacity, and by PAE the pre-application earnings. The optimization on the various sections of the budget constraint is as follows:

0%-50% of remaining earning capacity

As long as current earnings are below half of the remaining earning capacity, disability benefits are equal to the degree of disability, d_i, multiplied by unemployment benefits. Hence $DI(E_i) = d_i UI$ and $DI'(E_i) = 0$. Substitution into Equation (13) yields

$$\lambda_i = \frac{(T - E_i)w_i}{d_i UI + Tw_i} \tag{17}$$

$$E_i^* = T(1 - \lambda_i) - \frac{\lambda_i}{w_i} d_i UI \tag{18}$$

50%-100% of remaining earning capacity

Individuals earning between 50% and 100% of their remaining earnings capacity, receive disability benefits equal to 70% of their pre-application earnings, multiplied by their degree of disability: $DI(E_i) = 0.7 d_i PAE$. The amount of benefits they receive is independent of their number of working hours, and hence $DI'(E_i) = 0$. Solving Equation (13) for either the preference parameter or the optimal number of working hours gives:

$$\lambda_i = \frac{(T - E_i)w_i}{0.7 d_i PAE + Tw_i} \tag{19}$$

$$E_i^* = T(1 - \lambda_i) - \frac{\lambda_i}{w_i} 0.7 d_i PAE \tag{20}$$

Given that the amount of DI benefits is independent of the number of working hours, the implications are comparable to the 0%-50% of remaining earning capacity scenario: the optimal number of hours decreases, and the probability of working decreases.

100% of remaining earning capacity-65% of pre-application earnings

For those earning more than their remaining earning capacity, but less than 65% or their pre-application earnings, the amount of benefits decreases as current income increases: $DI(E_i) = 0.7d_iPAE - 0.7w_iE_i$. The derivative of the benefits is thus equal to $DI'(E_i) = -0.7w_i$. Solving for the preference parameter and the optimal number of hours gives:

$$\lambda_i = \frac{(T - E_i)0.3w_i}{0.7PAE + 0.3Tw_i} \tag{21}$$

$$E_i^* = T(1 - \lambda_i) - \frac{\lambda_i}{0.3w_i}0.7PAE \tag{22}$$

The negative derivative of benefits thus causes a strong reduction in the optimal number of working hours. For individuals who earn more than 65% of their pre-application earnings, all benefits are terminated and the optimization problem becomes identical to the pre-application problem.

A.4.2 Post-recovery

The last step of the model analyzes the response to recovery. Recovery can be modelled by increasing the hourly wage to its pre-application level and by lifting the hours constraint. Utility optimization will than give the optimal response to recovery. The obtained response should be interpreted as an upperbound as it assumes that all individuals fully recover and that all individuals can return to their pre-application job. Aspects such as distance to the labor market and partial recovery are thus not taken into account.

For individuals who have not been awarded disability benefits, lifting both the hours constraint and the wage constraint implies that their optimization problem is identical to the pre-application optimization problem. In contrast to this, individuals who have been awarded disability benefits do face a different optimization problem than they faced pre-application as their benefits are not terminated upon recovery. Their optimization becomes identical to their post-application optimization problem, which the exception that their hourly wage, and thus the slope of the budget constraint, increases and their hours constraint is lifted.