Back to work: Employment effects of tighter Disability Insurance eligibility in the Netherlands

Silvia Garcia Mandico[a], Pilar Garcia Gomez[a], Anne C. Gielen[a], Owen O'Donnell[a,b]

[a]*Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands*
[b]*School of Economics and Area Studies, University of Macedonia, Thessaloniki, Greece*

Abstract

In this paper we provide evidence on the impact of tightening the eligibility criteria for disability insurance (DI) on those who are already receiving disability benefits. To do so, we use administrative data on the universe of DI claimants to analyze a reform that took place in 2004, under which all DI recipients born after 1954 were reassessed under a stricter eligibility criteria. We exploit the exogenous variation in the implementation of the reform to explore its effect on several labor and health outcomes. First, we find that the stricter eligibility criteria increased transitions out of DI by 4.7 percentage points. We find that 50% of the disqualified claimants were employed two years after reassessment, whereas 22% were receiving unemployment benefits, and 38% remained out of the labor force. For those transitioning to employment, we find that the quality of the jobs they found marginally increased, and on average, the loss of income from DI was compensated with income from other sources (wages, unemployment benefits and social assistance benefits). Our results reveal that there are differences in the labor market behavior of individuals with different pre-existing conditions. We find that treated claimants suffering from mental disorders appear to have had sluggish transitions to employment, when compared to other medically verifiable conditions. Instead, claimants with musculoskeletal disorders had a higher employment response than the average claimant, suggesting a larger residual employment capacity. These differences highlight that a reform "one size fits all" may not be appropriate to foster employment among DI recipients, particularly when it comes to people suffering from mental health.

1. Introduction

Over the last two decades, the employment rates of people with disabilities have fallen relative to their counterparts without a disability. Among OECD countries, the relative employment for those with disabilities fell from 62% to 57% between 1995 to 2010 (OECD, 2010). The literature has reached consensus by attributing this trend to the generosity of the disability insurance (DI) programs, when paired with economic developments that widen the skills wage gap (Autor and Duggan 2003; Duggan and Imberman 2009; Burkhauser and Daly 2011). More recently, the low employment participation of those with disabilities has also been attributed to the liberalization of the DI program eligibility (Autor and Dugan 2006). The increased comprehensiveness of the scheme has attracted claimants suffering from subjective health conditions, such as from musculoskeletal and particularly mental disorders.[1] In 2010, the share of claimants suffering from mental disorders was close to 50% of the total DI claimants in the OECD countries (OECD, 2012). The share of new DI awards to people suffering from ill-health has experienced increasing trends in most industrialized countries, ranging from 27 to 47% in 2010 (OECD, 2012).

In the Netherlands, such developments led the government to implement a set of comprehensive reforms in the early 2000s. Those reforms were viewed as absolute game changes in the structure of the Dutch DI scheme, and are often considered as an example of a successful approach to curb the entry to DI. Together with decreasing the incentives to entering DI, these reforms were also found to have fostered employment among people with disabilities (Koning 2009, 2012). In this paper we provide evidence on the impact of one of those reforms on a much less explored population, i.e., those who are already receiving disability benefits. The reform under analysis entailed the

[1]The terminology 'subjective' arises from the fact that this type of conditions cannot be diagnosed through hard medical evidence; instead, the diagnosis is often solely based on the opinion of the health practitioner.

reassessment of all disability beneficiaries born after 1954 under a stricter eligibility criteria. The new eligibility rules were more flexible in considering which potential jobs recipients could still perform despite their disability. Instead, the medical criteria was left unchanged. The features of the reform provide a unique set up to analyze the residual employment capacity after disability benefit receipt. Our estimates apply to a very generalizable DI population, which is the working-age disability benefit recipients.

We explore the overall effect of the reform on the transition probabilities from DI to employment and non-employment for the average DI claimant aged in between 30 to 49 years old. We also focus on the heterogeneity of the employment and non-employment responses by cause of disability and age. Analyzing the heterogeneous effects by disability cause is particularly interesting for this reform, since the eligibility criteria was tightened independently of the medical condition. We then interpret the differences on DI exit rates across health conditions as the beliefs of the administration regarding the remaining work capacity associated to the different health conditions. The ratio between the probability of being in a particular state and the probability of leaving DI hinges on the different levels of error type II, moral hazard or labor market discrimination across diagnoses.

Generally, the literature exploring the employment effects of DI evaluates the work capacity of disabled individuals before disability benefit receipt (Bound (1989), Chen and van der Klaauw (2008), von Wachter, Manchester and Song (2011), Maestas, Mullen and Strand (2013), French and Song (2014)). Among those, the heterogeneity of employment effects per cause of disability has also received some attention. Watcher et al. (2011), Maestas, Mullen and Strand (2013) and French and Song (2014) provide evidence on diagnosis heterogeneity of the employment effects of DI for the US. Their findings suggest that DI applicants suffering from 'subjective' health conditions have the highest subsequent employment rates. However, benefit receipt may cause further labor detachment, by causing skill deterioration, social exclusion, depression or lack of self-esteem (see e.g. Vingard et al. 2004, Bryngelson et al. 2009; Svensson et al. 2010). Being so, the estimates on the initial work capacity of the disabled may rightfully differ to those on the work capacity after benefit receipt. Hence, the previous evidence may not be used to raise conclusions on the retained capacity to work of the DI population. Campolietti and Riddell (2011), Kostol and Mogstad (2013), and Borghans, Gielen and Luttmer (2014) offer some evidence on that matter. Campolietti and Riddell (2011) explore an increase in the earnings exemption in Canada, and find that it resulted in employment increases among existing DI recipients. The increase in labor force participation was not followed by higher exit rates from DI. Kostol and Mogstad (2013) use a regression discontinuity design to examine the effect of a return-to-work program in Norway. They find that the program induced an employment response only among younger individuals. Borghans, Gielen and Luttmer (2014) also use a regression discontinuity design, and examine the effects from a tighter eligibility criteria in the Netherlands, implemented in 1996. To date, Moore (2015) provides the only evidence on the retained employment capacity differentiating per cause of disability. In his analysis, he exploits a policy change in the US scheme in 1997, under which drugs and alcohol addictions no longer qualified for DI. This way, he is able to identify how work capacity changed with disability benefit receipt on the population whose first diagnosis was an alcohol or drug addiction[2]. His findings suggest that the employment response to DI withdrawal for individuals with medically verifiable conditions is lower than for those with musculoskeletal and mental disorders.

Because of the nationwide implementation of the reform, the identification of the impact of the reform falls in the category of age-cohort-time effect models. We correct for time and age effects by using a trend adjusted differences-in-differences with two counterfactual groups (Bell, Blundell, van Reenen 2009). First, a group of claimants the same age that our treated group in a period prior to the reform will be used to control for age effects. To adjust for cyclical dependencies, we also use information on the claimants who did not qualify for reassessment due to being 50 years of age or older in July 2004. We show the sizable employment response induced by the reform, despite the short term evaluation: out of an induced DI exit rate of 4.7 percentage points, 37% transitions to employment. In the longer term, the fraction of the DI exit rate transitioning to employment increases to 50%. Moore (2015) estimates an employment response after DI termination of 22 percentage points.

We contribute to the literature by exploring the quality of the employment of the disqualified DI claimants who do find a job, as well as the income effects of the reform. To our knowledge, we are the first to qualify the quality of the employment found after benefit receipt beyond the wage. We find that the reform did not induce major changes in the quality of the employment of the disabled, and if any, a small increase in the yearly net wage. Regarding the

[2]This implies that the benefit was granted on the basis of an alcohol or drug addiction. There was the possibility of a second diagnosis, as those diseases are often comorbid

income effects, we also find no major impacts of the reform. Our results seem to indicate that this reform fostered employment, without causing income or job quality losses among the disabled.

In terms of heterogeneity by health condition, our findings do not suggest that claimants suffering from subjective health conditions retain the highest employment capacity. In fact, claimants with musculoskeletal disorders seem to transition to employment just as much as those with neurological diseases, and certainly less than those with respiratory and cardiovascular diseases. Claimants suffering from mental disorders have a transition to employment well below that of people with medically verifiable conditions. Our results also suggest that the administration did not heavily discriminate between health conditions in terms of the strictness of the reassessment process.

Our paper contributes to the literature by analyzing the transitions to non-employment states induced by the reform. Our findings suggest that 22% of the DI exit rate transitions to the unemployment insurance (UI) program, and 38% moves to other non-employment states, such as general assistance and a temporary income program (TRI). Karlstrom, Palme and Svensson (2008) quantify the transitions from DI to employment, unemployment insurance (UI), sickness insurance (SI) and early retirement (ER). Using a difference-in-difference strategy, they exploit a policy reform in the Swedish DI program, which eliminated the special eligibility rules for the age groups 60-64. Their results do not find a significant transition to employment 2 to 3 years after the reform. The lack of employment effects seems to suggest that the group under analysis may generally have low retained emplyoment capacity, potentially due to their very old age. Instead, they find an increase in the transition to the UI program of a 25% of the DI exit rate, and a transition to the SI program of 12%, estimates very comparable to ours. Staubli (2011) follows a similar approach, exploiting this time a policy change in the Austrian DI scheme that tightened the eligibility criteria for 55-56 year old male DI recipients. They estimate transitions to UI to be in between 45-68% of the total DI exit rate, and transitions to SI to be of the order of 10%. Transitions to employment are of 32-45% of the total DI exit rate, a magnitude comparable to our findings. Borghans, Gielen and Luttmer (2014) provide evidence of spillover effects from DI to social assistance programs in the Netherlands. Their findings suggest that claimants exposed to a less generous DI benefit scheme have a probability of participation in other social assistance programs 7.8 percentage points higher. Their estimates of the employment response suggest that a boost in the DI exit rate of 10 percentage points was accompanied of an increase in the probability of being employed in 5 percentage points, so an effect very closely comparable to our long term estimates.

The extensive and detailed nature of the data allows us to explore the short-term health effects of the reform, through exploring changes in the probabilities of being hospitalized. We do not find any impact of the reform on this outcome, arguably due to the short time span of our observation.

By providing evidence on the effects of this reform, this paper contributes to the current policy debate on DI: should governments reduce the size of their disability programs by boosting the exit rate of the current claimants? Would this promote employment among the disabled? In this context, our short term findings would suggest so. However, our findings also suggest the need for a comprehensive set of measures to support the disqualified DI claimants. Transition to employment is not immediate, and subjecting the the disabled to an abrupt income shock may have negative consequences in their future income, quality of employment, or even health status. The Dutch government accommodated disqualified DI claimants with a temporary income, ensuring a smooth transition to the labor market. As a result, we do not observe large leakages to other social assistance programs, and instead observe a significant transition to employment. Rather than rushing to low paying jobs, our findings suggest that the DI claimants affected by the reform have used the time to find suitable jobs: thus the quality of the employment of the disqualified DI claimants has not been compromised. This is all the more important for individuals with mental disorders, which represent a large and very susceptible fraction of the DI claimants. Work integration is proven to be essential for the recovery for this subgroup, by providing social status, income security, a time structure, a sense of identity and achievement and a source of self-esteem, while also enabling social contact (OECD, 2012). However, as we see in our results, transition to employment is far from immediate for this subgroup, probably due to the stigma associated to such conditions.

The rest of the paper is organized as follows. Section 2 provides the institutional context and detailed information about the reform. Section 3 explains the potential impact of the reform, and section 4 describes the data. Section 5 presents the empirical strategy. Section 6 presents the results of the reform, first presenting the aggregate results, and then analyzing the heterogeneity in response by age and diagnosis. Section 7 concludes.

2. Institutional Context

2.1. The Dutch Disability Insurance

Before 2004, the Dutch disability insurance consisted of three benefit schemes: the WAO, targeted to workers between 30-65 years old, the Wajong, granted to the young disabled, and the WAZ scheme, which covered self-employed workers. Under the WAO scheme, workers were eligible for benefits if they were not able to resume their normal work schedule after one year of receiving sick pay. In that case, the social benefit administration (UWV) would assess their degree of disability, and grant them with disability benefits. To assess her degree of disability, the DI applicant first went through a medical assessment. If the physician considered that she still had some work capacity, the social insurance professionals would look at the set of jobs the applicant could perform despite her impairment. Then, the degree of disability was calculated as follows:

$$DD = 1 - \frac{\overline{\omega}}{\omega^p}$$

Where $\overline{\omega}$ is the average earnings of the 3 highest paying occupations the applicant could still perform despite her impairment, and ω^p are her labor earnings prior to the onset of the sickness. If DD fell below 15%, the applicant was not eligible for benefits; between 15% and 80% the recipient was classified into 6 different categories, and above 80% she was considered as fully disabled. Once the degree of disability was established, the corresponding benefit was calculated according to the following levels of wage replacement rates:

Degree of Disability	Wage replacement rate (%)
0-15%	0
15-25%	14%
25-35%	21%
35-45%	28%
45-55%	35%
55-65%	42%
65-80%	50.75%
80-100%	70%

The Dutch DI program prior to 2004 was particularly generous compared to other industrialized countries. First, a relatively wide range of medical conditions could qualify an applicant for benefits. This is so, since both work-related injuries or occupational diseases and non work-related sicknesses were considered during the medical assessment. Second, the replacement rates among fully disabled were more generous than in most other industrialized countries (OECD, 2009). Finally, the Dutch DI program offered a case-based eligibility, if we compare it with the US case, in which eligibility to the DI program is based on the applicant not being able to earn an income below the substantial gainful activity (SGA) threshold. This implies that the Dutch DI scheme contemplated the possibility of granting partial disability benefits for not fully disabled applicants. Partially disabled beneficiaries were allowed and encouraged to work, so that the benefit sometimes acted as a wage complement.

The outflow from DI consisted mostly of claimants finding employment paying at most 15% less of the pre-disability wage. After reporting a new job, DI claimants received benefits during a three month trial placement. Upon successful employment after this period, the benefits were terminated. The other source of DI outflow was the periodic reassessment of the stock of DI claimants. DI recipients underwent a first reassessment after one year of benefit receipt, then periodically every 5 years. These reassessments were often informal, under the form of questionnaires filled in by the claimant itself. Estimates show that they induced only 18 percent of the total outflow from DI (Work and Income Inspectorate (IWI) 2004).

2.2. The 2004 Reform

In October 2004, the existing stock of recipients born after the 1st of July 1954 was subjected to a reassessment of their application file under a stricter criteria. This reassessment process became one of the largest reforms in the Dutch DI system, lasting over 4 years and affecting 345.000 beneficiaries.

2.2.1. Change in the Eligibility Criteria

Broadly, the new eligibility criteria was more flexible when evaluating the suitable jobs for the applicants, which increased the range of jobs considered when calculating the degree of disability. The employment criteria was modified mainly by using the average wage of the highest paying 3 occupations for which 3 suitable jobs were available, instead of the highest paying 3 occupations for which 10 suitable jobs when calculating the degree of disability[3]. This change unambiguously resulted in a higher average wage from the suitable jobs considered, thus leading to a lower degree of disability. Another major change in the eligibility criteria was to re-calculate the pre-disability wage by truncating the maximal weekly hours worked to 38. This resulted in lower pre-disability wages, thus in a lower degree of disability. Other minor changers in the eligibility criteria were to consider jobs involving night shifts, full time jobs even for part-time applicants, and to match claimants with jobs specifically requiring the use of Dutch and computer skills even if the applicant didn't meet the requisites.

2.2.2. The process of reassessments

In April 2004, the congress approved a law according to which a reassessment reform would take place, starting on the 1st of July 2004. This reform would reassess the young DI recipients under the new eligibility criteria explained above, and had as a main goal to boost the outflow from DI. Because of strong opposition from unions and the lack of consensus about the exact criteria of the reassessments, the beginning of the reassessments was postponed until October 2004.

Initially, the plan was to reassess only these recipients born after July 1954, and to do so by groups of age, from young to old. Effectively, although younger recipients were reassessed more frequently at early stages of the reform, all eligible recipients of different age groups had a positive probability to be reassessed thorough the reform.

All eligible cases were re-assessed under the new, stricter criteria, often undergoing a medical assessment as well. At the end of 2005, the UWV published a very detailed report on the reassessment process[4]. Table 1 shows the outcomes of the reassessments. Overall, 32% of the reassessed recipients exited DI, i.e their degree of disability fell below 15%. Among initially fully disabled (degree of disability between 80-100%), a group that accounts for about 60% of all DI claimants (UWV 2006), 25% left DI. Around 8% experienced a decrease in their degree of disability, and over 60% had their degree of disability unchanged.

Table 1: Outcomes of the reassessments, detailed- End of 2005

	New disability level										
	< 15%	15 – 25%	25 – 35%	35 – 45%	45 – 55%	55 – 65%	65 – 80%	80 – 100%	unknown	total	obs.
Old Level											
15 – 25%	63.2	27.2	2.6	0.8	0.5	0.5	0.3	4.1	0.9	100%	8,189
25 – 35%	42.2	21.2	26.0	2.7	1.3	0.7	0.5	5.0	0.5	100%	7,260
35 – 45%	35.9	11.7	16.5	23.6	3.2	1.1	1.3	6.5	0.3	100%	5,319
45 – 55%	36.2	5.4	6.7	8.9	26.8	2.9	1.9	10.9	0.3	100%	5,470
55 – 65%	33.8	8.0	5.3	5.7	10.5	20.7	3.1	12.6	0.3	100%	2,978
65 – 80%	24.4	7.1	7.1	6.7	7.0	9.7	20.8	16.8	0.3	100%	2,340
80 – 100%	25.8	2.9	2.3	1.9	2.2	1.5	1.5	61.4	0.5	100%	55,895
total	32.3	7.7	5.7	3.9	4.0	2.3	1.9	42.0	0.5	100%	87,451

Source: UWV (2006)

At the end of 2005, the share of younger individuals reassessed was larger, so one would expect the outcomes from table 1 to become less extreme over time, under the assumption that older individuals may be more disabled. Additionally, in 2007, the newly formed government passed a bill following strong criticism of the reassessment policy. The new legislation stated that only those born on or after the 1st of July 1959 should be assessed under the stricter criteria. There were large consequences in the re-assessment process: out of the 115,000 recipients born in between July 1954 and July 1959, 53,000 had already been reassessed under the stricter regulation. This implied that a considerable number of recipients had to be reassessed a second time under the old more lenient criteria.[5]

[3]This implied then a reduction in the considered jobs from 3*10=30 employment alternatives, to 3*3=9 considered alternatives.

[4]Unfortunately, such detailed level of information was only disclosed once, at the end of 2005, so I cannot report the outcomes of the reassessments per old and new category of disability at a later stage of the reform.

[5]Particularly, 25,000 cases were revised (Staatsblad 2007, 324). That is: out of 53,000 recipients between 45 and 50 years old already reassessed, 20,000 were fully disabled under the strict criteria, hence not revised a second time. 2,000 individuals specially rejected the second reassessment. Note that the ages are taken as in July 2004.

Table 2 provides an overview of the final outcomes of the reassessment process: 20% of the individuals did leave DI because of the reassessment reform, and 12% experienced a decrease in their disability benefit while remaining on DI.

Table 2: Outcomes of the reassessments, total - Final

Outcome of the reassessment	Percentage
Higher DD	6%
Lower DD	12%
Unchanged DD	62%
Left DI (DD < 15%)	20%

Source: UWV (2009)

3. Data

We use administrative data maintained by Statistics Netherlands. The complete dataset employed for the analysis comes from different sources. First, demographic characteristics are obtained from municipal registries. These files contain information about the month and year of birth, gender, and place of birth and of residence for all residents of the Netherlands. Second, we use the social security records to obtain information on social assistance programs, particularly on DI, unemployment insurance (UI) and other social assistance[6]. Regarding DI, the social security records contain detailed information on the disability spells, main diagnosis, degree of disability and corresponding yearly payments. This information is available for the recipients in all schemes of the Dutch DI system: WAO, WAZ, Wajong and WIA (as of 2004). In this analysis, we will focus only on WAO recipients[7]. Regarding UI and other social assistance, we obtain and use from the social security records the spells and corresponding yearly payments. Finally, we obtain information on employment (and self-employment) from the tax authority records. Both the employment spells, yearly labor income and days worked per year are gathered and used in the analysis.

Each of the different sources contains an individual identification number (RIN-code) which allows us to merge datasets at the individual level. All data is organized in yearly records, and we make use of those from 2000 to 2006. As will become clear below, the empirical strategy followed implies that we will divide the sample in two (overlapping) periods: the comparison period, from 2000 to 2003, and the reference period, from 2003 to 2006. Individuals selected in the comparison period will be treated as distinct to those in the reference period, even though their identification number indicates that they are the same person. For simplicity and to differentiate individuals from different periods, we construct identification variable on the basis of the RIN-code and the selected period, which we name *ID*. This approach enables us to use the individuals from the comparison period as counterfactuals for the individuals in the reference period.

We limit and modify the data employed in several directions. First, we re-organize the datasets on a monthly basis. Since there is precise information on the employment and social assistance spells of each individual, we construct binary variables specifying the status of the individual in every month, for every year[8]. Second, we select individuals who are in the DI status at the beginning of the period of analysis (2000 for those in the comparison period, and 2003 for those in the reference period), and follow them throughout, so that we effectively work with a balanced panel. The only source of attrition in our sample is due to death, and to individuals leaving both employment and any social assistance program. In the reference period sample, we leave out all individuals younger than 30 years old in July 2004[9]: we obtain a sample of 103.674 individuals eligible for the reform (i.e aged in between 30 to 49 in July 2004) that we will follow on a monthly basis over time.

[6]Other social assistance is defined as the sum of general assistance, sick pay, and a set of small and very specific social assistance programs.

[7]WIA recipients are omitted because they were not part of the reassessment wave. The self-employed may differ in characteristics to the employed, so WAZ recipients are also omitted. Finally, Wajong recipients have never worked before, making their labor participation different, and are omitted as well. The WAO scheme, in any case, encompasses 80% of DI recipients in the Netherlands.

[8]That is, status indicator variables taking value 1 for an individual in the sample if he or she has been in this status for at least one day of that month.

[9]Analogously, we leave out all individuals younger than 30 years old in July 2001 in the comparison period sample

Table 3 shows the sample descriptive statistics by age group, for time periods before and after the reform. Because the focus is on the reference period sample, the period before the reform is defined from January 2003 to September 2004, and the after reform period is defined from October 2004 to December 2006. The first panel of the table shows the average proportion of the individuals during the pre- and post-reform periods in each of the different possible states. Disability (Total) measures the average proportion of individuals on DI in the sample. Disability only and employment and disability show two of the different possibilities for those being on DI: being on DI only, or being on DI and employed at the sample time[10]. Employment, unemployment and self-employment show the average proportions of the sample in employment, in UI and in self-employment. New Disability spell captures the probability of re-entering DI after January 2003. Others corresponds to all those individuals who cannot be classified in any of the previous categories.[11] We have separated also those individuals who died in the sample. We see some sizable differences in the before-after DI exit rate, which are particularly sizable for these eligible to the reform. This highlights again magnitude of the impact that the reassessments had on the outflow from DI.

The second panel gives some general characteristics on the individuals in the sample, such as the annual income (expressed as the sum of the annual incomes from work, social assistance and self-employment), gender, average age and percentage of fully disabled. We also explore the average time claimants had spent on DI at the beginning of the observation (January 2003), and divide them into those who left DI during the pre- or post-reform period, and those who didn't. We observe that those claimants leaving DI before the reform had spent about half the time on DI than those who remained on DI. This ratio increases slightly after the reform. It could be due to the fact that before the reform, the DI exit rate was composed by these recovering and finding employment, which arguably could also have spent less time on DI. Instead, the reform boosted the exit rate, by reassessing and disqualifying individuals who would not have left hadn't there been a reform.

Finally, the last panel offers an overview of the main diagnoses of the claimants in the sample, as reported in 2003. We have selected the groups of diagnoses with an incidence of at least 1% of the sample population. The detailed conditions that correspond to general diagnosis can be found in the appendix. Younger claimants tend to suffer the most from mental disorders; Instead, the incidence of musculoskeletal disorders, second largest diagnosis for the treated individuals, appears to increase with age.

[10]Note that there are other combinations possible, such as being on DI and receiving unemployment benefits, or other social assistance. Those combinations represent a negligible proportion of the sample states, so are included in the category disability only.

[11]Note that by this definition, they could be earning other social assistance only. We do not make a distinct category, since this state represents marginally a zero percent of the states.

Table 3: Descriptive statistics by age before and after the 2004 reform.

	Eligible for reassessment: ages 30-49		Non-eligible for reassessment: ages 50-53		Non-eligible for reassessment: ages 54-55	
	Before Reform	After Reform	Before Reform	After Reform	Before Reform	After Reform
States (%)						
Disability (TOTAL)	93.39	79.15	96.94	92.24	97.59	93.94
Disability Only	63.97	57.16	65.18	65.86	66.54	68.41
Employment and Disability	29.29	21.92	31.77	26.32	30.97	25.48
Employment	3.86	11.06	1.95	4.48	1.49	3.22
Unemployment	0.87	2.57	0.31	0.74	0.24	0.66
Self-Employment	0.27	1.70	0.18	1.14	0.67	2.19
New Disability spell	0.22	0.87	0.10	0.28	0.07	0.19
Other	1.38	4.98	0.64	1.98	0.55	1.70
Sample Characteristics						
Annual Income	17,050	16,453	19,784	19,306	20,283	19,705
Time on DI (in January 2003, in years)						
In DI	6.16	6.58	9.27	9.48	10.85	11.07
Not on DI	3.00	3.53	4.64	4.83	6.15	6.37
Average age (July 2004)	42		52		54.50	
Female (%)	57.48		45.56		42.53	
Fully disabled (%)	64.50		65.76		66.58	
Death (%)	0.42	1.34	0.80	2.45	0.86	2.72
Main diagnosis(%)						
General diagnosis	9.98		12.22		12.55	
Cardiovascular diseases	2.37		4.89		5.47	
Musculoskeletal disorders	29.96		32.86		33.53	
Neurological condition	6.08		5.43		4.99	
Mental Disorders	41.14		34.20		32.07	
Respiratory	1.43		2.02		2.20	
Digestive	2.33		2.17		2.31	
Health Outcomes						
Hospitalizations (%)	2.38	2.39	2.62	2.65	5.49	5.70
Emergency hospitalizations (%)	0.96	0.94	1.06	1.06	2.32	2.40
Nights of hospitalization	7.83	7.34	8.38	7.22	8.88	7.83
Observations	5,140,674	6,609,438	1,938,258	2,492,046	576,135	740,745

Note: This table presents the sample statistics for different groups of age, for periods before and after the reform. The individuals are categorized in age groups on the basis of their age in July 2004, as was determined by the reform. Then, the first column shows the sample means for the claimants eligible for the reform (treated group), and the second and third column shows sample means for claimants above the age requirement to be part of the reform. The pre-reform period is defined from January 2003-September 2004, and the post-reform period from October 2004-December 2006. Because we stop the sample period in 2006, we also include in the treatment group also those individuals aged in between 45 to 49.

4. Empirical Approach

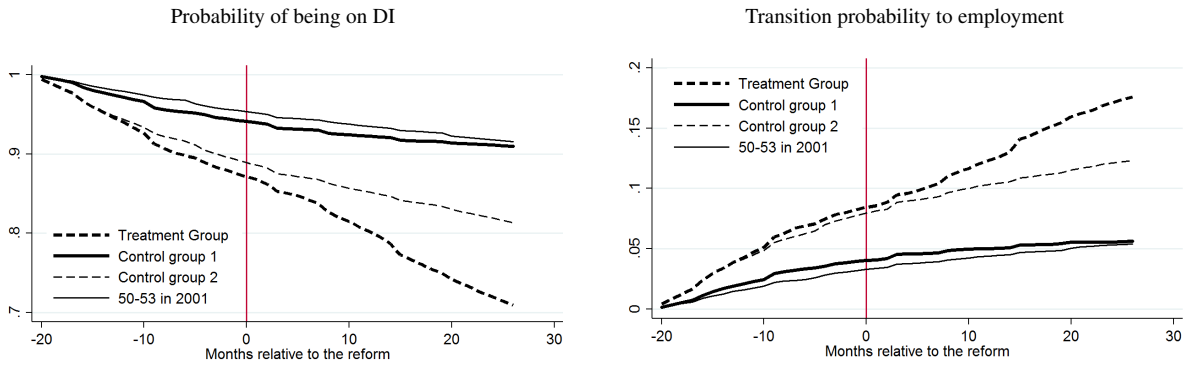
We use a differences-in-differences framework to identify the average impact of the reform on the recipients eligible for reassessment. One comparison group could be those aged over 50 in July 2004, hence not eligible for reassessment. This strategy can be problematic, particularly as the age gap between the treated and comparison claimants widens. One could expect older DI beneficiaries to have a lower probability of return to work than younger recipients, invalidating the common trend assumption in the pre-reform period between the two groups. Table 3 shows significant differences in the ratios for the different states across age groups in the period prior to the reform, which could expose age-related trend differences.

To avoid age-induced trend differences, an alternative comparison group could be those DI claimants of the same age as the treated beneficiaries in the comparison period, a period selected prior to the reform. In the presence of time specific economy-wide changes, however, the common trend assumption required for this approach to identify the effect of the reform is also doomed to fail. As previously introduced, the comparison period is set to be from January 2000 to December 2003, and the reference period from January 2003 to December 2006. The choice of the comparison period is not trivial, as it requires (1) similar macroeconomic conditions with the reference period, and (2)

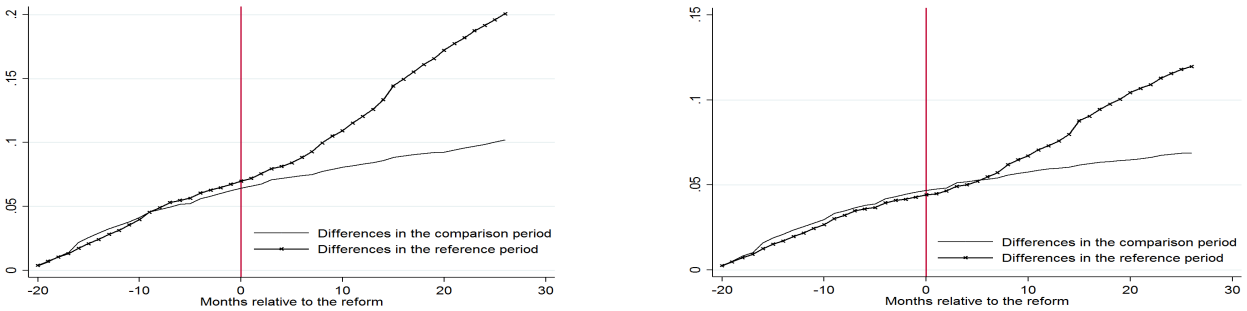
no major reforms in the social insurance system or the labor market that would affect the trends of the DI claimants in the comparison period.

Figure 1: Trends and time adjusted trends in the DI exit rate and transition probability to employment for the treated and potential control groups

Panel A: Trends in the DI exit rate and transition probability to employment



Panel B: Age-differential trends in the DI exit rate and transition probability to employment



Note: This figure displays the sample mean of those individuals in a particular state, interpreted as probabilities. The evolution of DI exit rate and transition probabilities are shown for the treatment group, and two control groups. The treatment group is defined as those aged in between 30 to 49 in July 2004. Its trends are shown for the reference period, i.e. January 2003 (Month -21) to December 2006 (Month 26). The time of the reform, October 2004, is normalized at zero. The control group 1 are those claimants aged in between 50 to 53 in July 2004, and are followed from January 2003 to December 2006. The control group 2 are those aged in between 30 to 49 in July 2001, and are followed from January 2000 to December, which constitutes the comparison period. Those DI claimants aged in between 50 to 53 in July 2001, and followed in the comparison period are also added. Panel B displays the trend differential between young and old claimants in both the reference and comparison periods. For the reference period, this is constructed by taking the absolute difference in the sample means between those aged in between 30 to 49 in July 2004, and those aged in between 50 to 53 in July 2004, for a particular state. For the comparison period, mean differences are taken for those aged in between 30 to 49 in July 2001 and those aged in between 50 to 53 in July 2001. Again, the time of the reform, October 2004, is normalized at zero.

Panel A in figure 1 shows the monthly trends for the different outcomes for the treated DI claimants (thick dashed line), and for the two comparison groups previously described: those aged over 50 in July 2004 (control group 1, thick solid line), and those the same age as the treated in the comparison time period (control group 2, thin dashed line). The x-axis shows the months relative to the reform (October 2004), which is set at the origin. The reference period starts 21 months before the reform, in January 2003, and finishes 26 months after, that is December 2006. Analogously, for the comparison period, month -21 corresponds to January 2000, and month 26 corresponds to December 2003. That way, the graph allows the comparison between groups in the reference period and in the comparison period. We include also those claimants aged in between 50 to 53 in July 2001, and followed during the comparison period (50-53 in 2001, thin solid line).

The trends between the treatment group and control group 2 match closely at the beginning of the observation periods for the exit rate from DI and the transition probability to employment. Instead, control group 1 presents

flatter trends in the exit rate from DI and transition probability to employment. One reason could be different costs of working for young and old DI claimants, due to differences in both groups' diagnosis composition. As table 3 showed, younger individuals tend to suffer more often from mental disorders, which could be associated to a higher recovery rate from DI and subsequent transition to employment (Autor and Duggan (2006), Von Wachter et al. (2011)). Alternatively, older individuals also have generally longer DI spells, possibly causing skills depreciation and labor market detachment (Vingard et al. (2004), Bryngelson et al. (2009), Svensson et al. (2010)). The trends of the DI exit rate and transition to employment between the treatment and control group 2 start diverging ten months prior to the reform (i.e January 2004). We cannot rule out the presence of anticipation effects from the reform; since the reform was announced prior to its implementation date, DI claimants expecting a negative outcome from the reassessment may have anticipated the reform by seeking a job, for example.[12] However, when comparing the trends of control group 1 with those of the same age in the comparison period, we also note a gap arising ten months prior to the reform. It seems to be that rather than due to anticipation effects, which should not arise for control group 1, the gap is caused by time-specific factors. The question that arises form the previous discussion is, wouldn't there be time effects, would then control group 2 be a suitable counterfactual for our treated group? To answer this question, Panel B in figure 1 shows the absolute difference in the trends between control group 1 and the treated group, and control group 2 and those aged 50 to 53 in 2001. The graph shows then the trend differentials between young and old in the reference period (crossed line) and comparison period (solid line). Note that the age differential outcome is no longer linearly dependent on time, so we can now assess the suitability of the comparison period claimants as a counterfactual for the reference period claimants. We observe that the age-differential trends in both the comparison and reference periods follow a very close pre-reform trajectory, giving support to the validity of a common trend assumption between them. This gives us confidence that the impact of the reform on the DI exit rate can be identified using the age-differential DI exit rate.

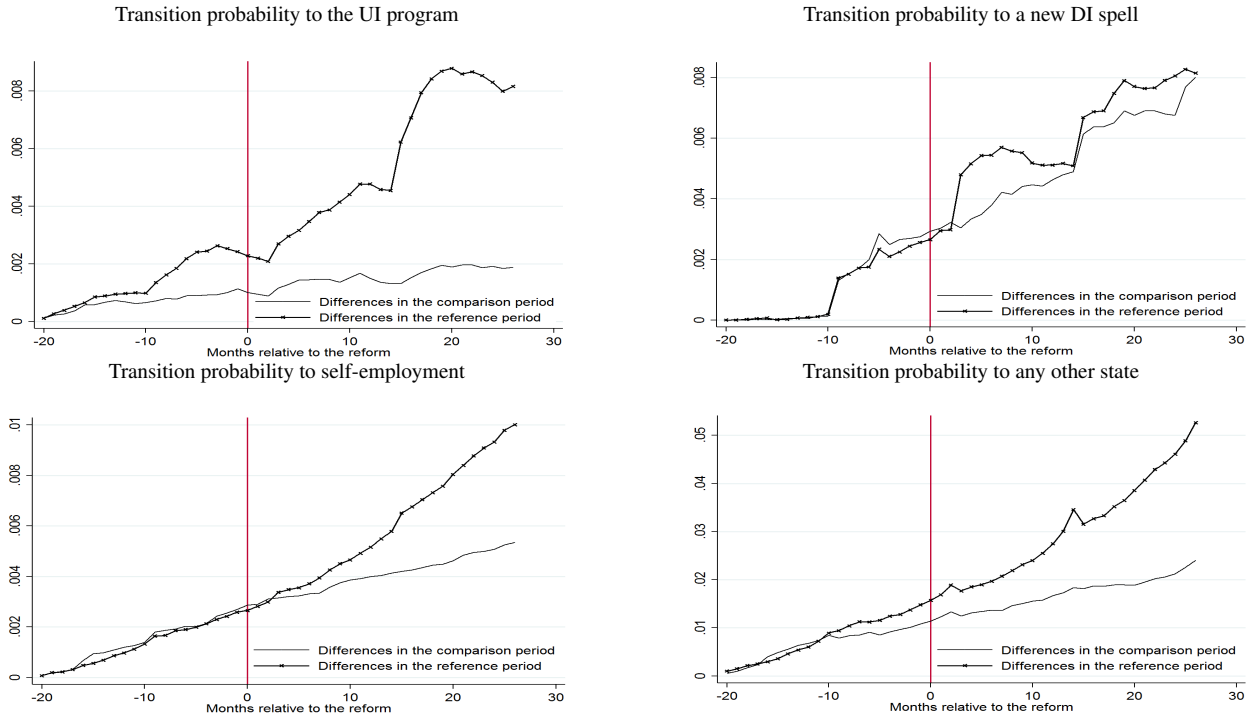
To emulate this approach in our empirical analysis, we use a time trend adjusted differences-in-differences model, first presented by Bell, Blundell and Van Reenen (1999). In their analysis, they aim at identifying the impact of a labor market reform for the young unemployed in the UK (the "New Deal" reform), implemented at the national level. Similarly to the identification issue occurring in our set up, the older unemployed in their analysis may have very different productive characteristics to the treated unemployed, and thus may not be satisfy the common trends assumption. Their approach is then to examine how the older group have responded differentially to the state of the cycle compared to the treatment group in similar periods of the business cycle to control for this effect. To do so, they exploit the pre-programme behavior of the New Deal target group and groups of older age.

Both Bell et al. (1999) and our approaches fall inevitably in the identifiability problem of age-period-cohort models: because of the exact linear dependency among these three factors, their simultaneous identification is impossible. For our case, this implies that although we explore trend differentials across groups of age, thus accounting for time and age effects, the identification of the treatment effect on this outcome will not be consistent in the presence of cohort effects. From the trends in the probability of being on DI and transitioning to employment shown panel B figure 1, we do not expect cohort effects to play a major role for these outcomes. However, this may not be the case for all transition probabilities analyzed.

In fact, figure 2 shows the age differential trends in the transition probabilities to non-employment states and self-employment, both in the reference and the comparison periods. The non differential trends can be found in figure 10 in the appendix. The age-differential trends in the transition from DI to self-employment follow a close pre-reform pattern. Whereas graphically one can see a clear impact of the reform on the transition probability to self-employment, the trend in the transition probability to a new DI spell is more erratic, hence makes it difficult to graphically quantify any impact from the reform. Both the age-differential trends of the transition probability to the UI program and to other states show different pre-reform trends between the reference and comparison periods. Particularly, the differential trends in the reference period experience a steeper increasing trend already in the pre-reform period, which may lead to a substantial overestimation of the impact of the reform if using the approach of Bell et al. (1999). This will be taken into account in the interpretation of the subsequent analysis.

[12]The reform was announced in May 2003, with the entry of the new cabinet. However, it was not until April 2004 when the plan of the reassessments were formalized.

Figure 2: Age-differential trends in the transition probabilities to non-employment states for the treated and potential control groups



Note: This figure displays the trend differential between young and old claimants in both the reference and comparison periods. For the reference period, this is constructed by taking the absolute difference in the sample means between those aged in between 30 to 49 in July 2004, and those aged in between 50 to 53 in July 2004, for a particular state. For the comparison period, mean differences are taken for those aged in between 30 to 49 in July 2001 and those aged in between 50 to 53 in July 2001. Again, the time of the reform, October 2004, is normalized at zero.

4.1. Derivation of the empirical strategy

Let the true relationship between the probability of being in a particular state and the effect of the reform be:

$$y_{igpt}^{state} = \delta D_g T_t + u_{igpt} \tag{1}$$

$$= \delta D_g T_t + \mu_{it} + \epsilon_{igpt} \tag{2}$$

Where y_{igpt}^{state} is a binary indicator of being in a particular state for a claimant i, in age group g, at month p relative to the (pseudo) reform and calendar time t. T_t is a time dummy that takes value 1 at the onset of the reassessment process, i.e October 2004. D_g takes value 1 for young DI claimants from both the reference and comparison periods ($g = Y$), and 0 for older claimants ($g = O$). The unobservables follow a standard approach, and contain an individual time trend μ_{it}, and an idiosyncratic error term ϵ_{it} (iid over i, g, p and t).

To illustrate the consistent estimation of the treatment effect, we derive the estimators from different identification strategies in a discrete time set up. Let $\{t_0; t_1\} = \{2003; 2006\}$, and $\{t_{0*}; t_{1*}\} = \{2000; 2003\}$, so that the reference and comparison periods are spanned, respectively. An approach previously discussed is the estimation of the treatment effect from (1) with a difference in difference estimator, using claimants the same age in different time periods. This is,

$$\hat{\delta}^{DID} = (y_{Ypt_1}^{state} - y_{Ypt_0}^{state}) - (y_{Ypt_{1*}}^{state} - y_{Ypt_{0*}}^{state})$$

Where $y_{Ypt_j}^{state} = \mathbb{E}[\, y_{igpt}^{state} \mid t = t^j,\ p\ g = \{Y\}\,]$, and $\mathbb{E}[.]$ is the expectations operator. What is required for $\hat{\delta}^{DID}$ to be a consistent estimator of δ? We need

$$\mathbb{E}[\, u_{igpt} \mid t = t^j,\ g = \{Y\}\, p\,] = \mu + \kappa_{Yp} \tag{3}$$

Thus, the identification of the treatment effect allows the structure of the error term to have a time invariant individual component μ, and a time relative to the (pseudo) reform effect that varies across age groups. Note that because only individuals of the same age group g are employed in this strategy, the time relative to the reform effect will be the same for all individuals. What this identifying assumption implies is that only the distance between the time at observation and the time of the reform (or pseudo-reform, in case of the comparison period) should have an effect on the outcome of interest. Time period specific developments should not affect the transition probabilities. However, we saw from the graphical analysis that generally, time specific economy-wide changes played a role in the trends of our outcomes of interest. This implies that $\hat{\delta}^{DID}$ may not be a consistent estimator of δ, to the extent that it would capture not only the impact of the reform, but would confound it with time specific economy-wide changes.

In order to control for time specific effects, we follow Bell et al. (1999) approach of a trend adjusted diff-in-diff. As previously explained, this estimation procedure aims at identifying the treatment effect by using the differential evolution between young and old claimants in the comparison period as a counterfactual for the differential evolution between treated and older claimants in the reference period. By focusing on age-differential trends instead of the trends, we can account for general time effects. To see this, we derive the time-trend adjusted difference-in-difference estimator in the two period example:

$$\hat{\delta}^{TADID} = [(y^{state}_{Ypt_1} - y^{state}_{Ypt_0}) - (y^{state}_{Ypt_{1*}} - y^{state}_{Ypt_{0*}})] - [(y^{state}_{Opt_1} - y^{state}_{Opt_0}) - (y^{state}_{Opt_{1*}} - y^{state}_{Opt_{0*}})]$$
$$\text{where } y^{state}_{gpt_j} = \mathbb{E}[\, y^{state}_{igpt} \mid t = t^j,\ g = \{Y; O\},\ p\,]$$

In this new specification, $\hat{\delta}^{TADID}$ is a consistent estimator of δ if

$$\mathbb{E}[\, u_{igpt} \mid t = t^j,\ g = \{Y; O\}\, p\,] = \mu + \kappa_{gp} + \gamma_{t^j} \tag{4}$$

As previously, the identification of the treatment effect allows for an individual time invariant term and a time relative to the reform effect that is allowed to differ across age groups. Additionally, the trend-adjusted diff-in-diff approach allows for calendar time effects γ_{t^j}.

The presence of cohort effects would introduce a non-linearity between calendar time t and age group g, which cannot be controlled for using this approach. In words, cohort effects imply that age-differential trends differ across periods prior to the reform, invalidating the trend-adjusted diff-in-diff approach. In light of the previous graphical evidence, we argue against the presence of cohort effects for the probability of being on DI, to transition to employment, self-employment and re-entering on DI. Instead, we are cautious when drawing conclusions on the impact of the reform on the transition probability to UI and to other states, which may be overestimated using this approach.

Note that none of the specifications presented here control for the presence of anticipation effects to the reform. In fact, we know that the reform was formalized in March 2004, so that DI claimants could anticipate by transitioning from DI to other states before the onset of the reassessments. Being so, our estimates should be interpreted as a low bound for the reform in the following sections.

5. Results

The results are presented in six subsections. The first subsection examines the transitions out of disability, and transition probabilities to employment and non-employment states. The second subsection discusses the effects of the reform on the quality of employment. The third and fourth subsections quantify the income effects and the short-term health effects. The fifth subsection separately analyzes the impact of the reform per diagnoses having led to DI award. Finally, the sixth subsection discusses the heterogeneous effects of the reform by time in DI.

5.1. What is the effect of the reform on labor market outcomes?

Let the empirical counterpart of the differences-in-differences approach previously described, after imposing the common trend assumption (3), be:

$$y^{state}_{igpt} = \delta D_g T_t + \mu_i + \sum_{p=-20}^{26} m_p + \epsilon_{igpt} \tag{5}$$

Furthermore, the empirical counterpart of the trend-adjusted differences-in-differences approach after imposing the no-cohort effect assumption (4) is:

$$y_{igpt}^{state} = \delta D_g T_t + \sum_{p=-20}^{26} m_p + D_g \sum_{p=-20}^{26} m_p + \mu_i + \gamma_t + \epsilon_{igpt} \tag{6}$$

Both specifications are indexed over individuals i, defined at the *ID* level, groups of age g, calendar time t, and adjusted time since the reform p. In specification (5) we use only young individuals, so that $g = Y$ for all i, whereas in specification (6) both young and old individuals are used, so that $g = [Y; O]$. We model the age group per time since the reform effects, κ_{gp}, by $\sum_{p=-20}^{26} m_p$ after imposing the common trend assumption. After imposing the no cohort effects assumption, κ_{gp} is modeled by $\sum_{p=-20}^{26} m_p + D_g \sum_{p=-20}^{26} m_p$. Both specifications are estimated using fixed effects models: individual fixed effects for (5), and individual and time fixed effects for (6). We cluster standard errors clustered at the *ID* level (Donald and Lang (2001)), since the treatment is defined at the individual level.

Table 4: Estimates of the impact of the reform on the DI exit rate and transition probabilities

	Difference-in-difference	Trend adjusted difference-in-difference
Probability of being		
on Disability	-5.01***	-4.66***
	(0.07)	(0.11)
in Employment and Disability	-1.25***	-0.54***
	(0.08)	(0.13)
on Disability Only	-3.77***	-4.15***
	(0.08)	(0.13)
Employment	1.62***	1.73***
	(0.06)	(0.08)
Unemployment	1.25***	1.04***
	(0.02)	(0.03)
Self-Employment	0.14***	0.20***
	(0.02)	(0.02)
New Disability spell	0.05	0.08
	(0.04)	(0.06)
Other	1.94***	1.77***
	(0.08)	(0.05)
Observations	467,558	653,933

Note: This table shows the estimates of the difference-in-difference specification from equation (5) *and the triple difference-in-difference specified in equation* (6) *for the average treated DI claimant. The estimates have been multiplied by 100, so that they are readily interpretable as percentage points variations. Significance stars indicate significance of the estimate at the 1 percent level ***, 5 percent level **, 10 percent level *. Standard errors are in parentheses, clustered at the individual level. The observations correspond to the number of individuals in the first period observed.*

Table 4 shows the baseline estimates for the entire sample, resulting from the estimation of specifications (5) and (6). The first panel shows the estimated impact of the reform on the probability of being on DI, which we further decompose into the probability of being both employed and on DI and on DI only. The face value of the point estimates obtained from both models do not differ greatly, albeit being statistically different due to the small size of the standard errors. The results show that the DI exit rate was boosted by 4.66 and 5.01 percentage points after the reform for the trend-adjusted diff-in-diff and the diff-in-diff model, respectively. The diff-in-diff specification produces larger estimates, in line with the fact that they may be confounding the cyclical component, as noted in graphical analysis. Both models coincide in the reform having a stronger impact on the probability of leaving DI alone than on the probability of leaving both DI and employment. This can be given several interpretations. First, it is possible that the reassessments differed between those who were employed and those who were not. For instance, if the administration perceived employed DI claimants as signaling their willingness to work, they could have been subject to a lower short-run frequency or lower strictness of the reassessments. The second reason could be that there was reshuffling between the categories of DI only and DI and employment: it could be that the exit rate from those on DI and employment is offset by an inflow from those individuals initially on DI only, who remain on DI and find a

job after the reform. This transition could be a response to the reform having lowered the disability degree for those initially on DI alone.[13]

The second panel of figure 4 shows the transition probabilities to employment and non-employment states. Again, the magnitude of the estimates from the two models is similar. The transition probability to employment experienced a significant increase in 1.62 to 1.73 percentage points for the diff-in-diff and trend adjusted diff-in-diff models, respectively. The difference in estimates obtained with he two models is not statistically significant. This is an interesting result in itself, as it would suggest that calendar time effects did not play a major role: DI claimants were able to transition to employment in a similar fashion both in the reference and comparison periods. This is surprising, given the set of reforms implemented in the DI scheme in between the two periods. In 2002, the gatekeeper protocol was implemented, dramatically increasing the cost for employers to send their employees to DI. One of the perils of reducing the incentives of employers was that it could have negatively impacted the hiring possibilities of employees with a DI history. By not observing a different impact from time trends in the probability to transition to employment for cohorts taken before and after the gatekeeper protocol, our results suggest that the probability of becoming employed for these exiting DI does not seem to have been affected by the protocol.

The estimate on the probability to receiving unemployment benefits shows an increase in between 1.04 to 1.25 percentage points, depending on the model chosen. The estimates for the transitions probabilities to re-entering on DI with a new disability spell are not statistically significant. The estimates of the probability of becoming self-employed are significant, and albeit being very small we note that the impact of the reform represents a sizable boost given the fraction of the sample in this state (0.27 %, see table 3). There seems to be a sizable transition to the residual category other. Besides including individuals not receiving any sort of income support, this category also includes individuals receiving general assistance, and payments from the *Temporary Income Program* (TRI, *Tijdelijke regeling inkomensgevolgen*). Under the TRI, reassessed individuals were entitled to an allowance which ensured their pre-reassessment income for a maximum period of 12 months.[14] Similar evidence on the employment and non-employment effects were found in Borghans et al. (2014). In what follows, and to ensure minimal overestimation of the impact of the reform, our preferred model will be the trend-adjusted difference-in-difference.

To examine the fraction of the DI exit rate of the different transition probabilities, we define:

$$\hat{f}^{state} = \frac{|\hat{\delta}^{state}|}{|\hat{\delta}^{DI}|}$$

Where $\hat{\delta}^{state}$ is the estimated impact of the reform for a particular state. Table 5 shows that the largest fraction of the DI exit rate transitions both to employment and other non-employment states: 37% and 38% of the DI exit rate, respectively. Regarding unemployment and self-employment, they consist in a 22% and 3% of the DI exit rate, respectively.

[13] Including a dummy indicating a decrease in the degree of disability in specifications (5) and (6) retains some explanatory power in the probability of being employed and on disability.

[14] Unfortunately, our data does not allow us to identify the individuals in this program, so those only those receiving the TRI but not in any of the other specified states will be captured in the state *other*.

Table 5: Estimated fraction of the main transition probabilities over the total DI exit rate

	$\hat{f}^{state}$
Outcome variables	
Employment	0.37
Unemployment	0.22
Self-Employment	0.03
Re-entry on DI	0
Other	0.38
Observations	653,933

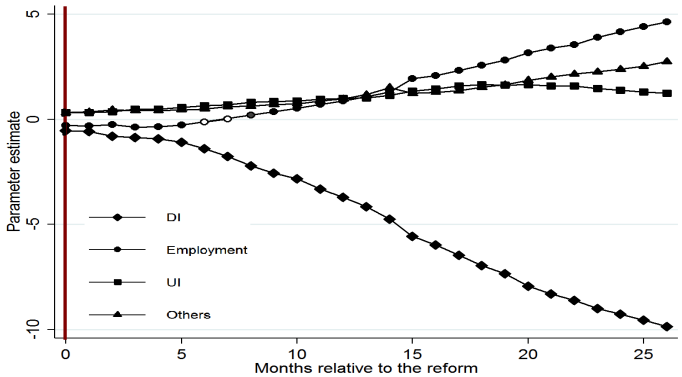
Note: This table shows the estimated fraction of the total DI exit rate for each state, $\hat{f}^{state} = \frac{\hat{\delta}^{state}}{\delta DI}$. $\hat{\delta}^{state}$ *is the estimated impact of the reform for a particular state from specified in equation* (6).

The previous approach, which estimates a static impact of the reform, imposes the underlying assumption that the full impact of the reform has been reached at the onset of the reform (Mora & Reggio, 2012). However, the process of transitioning to employment may not be immediate after reassessment, but instead consist of interesting dynamics which cannot be observed with the previous approach. We estimate a fully flexible model of the impact of the reform with time, allowing our estimates of the DI exit rate and transition outcomes to vary over time after the onset of the reassessments. Let the set R contain the post-reform time periods. The specification (6) is then rewritten as

$$y_{igpt}^{state} = \delta D_g \sum_{t \in R} \gamma_t + \sum_{p=-20}^{26} m_p + D_g \sum_{p=-20}^{26} m_p + \mu_i + \gamma_t + \epsilon_{igpt} \tag{7}$$

Figure 3 shows the estimates from the flexible specification (7) for the DI exit rate and the main transition states[15]. In the horizontal axis, 0 corresponds to October 2004, i.e the onset of the reassessment process.

Figure 3: Flexible estimates of the impact of the reform on the DI exit rate and main transition probabilities



Note: This figure shows the flexible estimates from specification (7)*, with clustered standard errors at the individual level. The horizontal axis, 0 corresponds to October 2004, the onset of the reassessment process, also marked by the red vertical line. The estimates for each month after the onset of the reform are then plotted. The color of the markers indicates the significance level of the estimates. Black markers indicate a significance of the estimate at the 1% confidence level, dark gray at the 5% confidence level, light gray at the 10% confidence level, and white indicates non-significance. The parameter estimates have been multiplied by 100, so that they are readily interpretable as percentage point variations.*

We see that the sizable impact of the reform on the exit from DI starts about 5 months after the onset of the reform. This is plausible, given that UWV sources claim that the reassessment process effectively started a few months after

[15]For ease of exposition, we focus on the states more largely affected by the reform: the transition probabilities to self-employment and re-entering DI are thus excluded from figure 3.

its official onset, due to the political opposition to this reform (UWV, 2007). By the end of our observation period, the effect on the DI exit rate reaches a 10 percentage point increase. The transition to employment also seems to have experienced a sluggish start, but reaches an increase of up to 5 percentage points by the end of the observation period. This means that the ratio of the transition to employment over the DI exit rate increases to 50% when we account for flexible estimates. As suspected, the transition to employment is not immediate after leaving DI: one could expect this ratio to become even larger in the longer run. Regarding the dynamics of the transition to employment, we note that during the first year after onset of the reassessments individuals seems to transition more frequently both to unemployment and other states than to employment. This is probably due to the fact that reassessed DI recipients were allowed to retain their entitled UI benefits at the time of entering DI during the entire time they were in DI. Once reassessed and disqualified from DI, transitioning to UI seems then a natural decision. For reassessed claimants who were not entitled to UI, or who were entitled to 6 month or less of UI benefits, they could qualify for the TRI program. Both social programs being particularly accessible for ineligible DI claimants, can explain the short run boost in the transition to UI and other non-employment states. Still, we note some differences in the evolution of the estimates of the transition to unemployment and to other states. Whereas the latter follows a rather linear evolution, the transition towards unemployment peaks at the 20th month after the reform, and then decreases, reaching a 1.23 percentage point increase at the end of the period.[16] This decrease is not accompanied by a boost in the transition probability to other non-employment states, but in the transition to employment only. It seems thus that disqualified DI claimants received their entitled UI benefits, and subsequently transitioned to employment rather than to leaving the labor force altogether.

Our results so far show a rather optimistic outlook of the reform. Already in the short run, we observe that half of the DI exit rate induced by the reform moves to employment. Particularly, this fraction seems to increase over time, giving the idea that it could lead to further transitions to employment in the longer run. Instead, there is a decreasing fraction of the transitions to non-employment states, mainly due to a fall of the impact on the transition to unemployment. This fall is followed by a transition to employment rather than to non-employment states, which could indicate the use of the time in UI to search for a job. Regarding the transition to the residual non-employment states, despite the termination of the TRI benefits, we do not observe a fall of the impact of the reform on this transition, but rather a constant share of disqualified DI claimants moving to this state. One possibility would be that the individuals that self-select themselves into transitioning to residual non-employment states have a low income elasticity: the income drop from terminated TRI benefits does not foster a transition to employment, nor increases the exit rate from the other non-employment states. If this is an indication of their true incapacity to work, we could expect a re-entry do DI, which we do not observe, or a worsening of their health status, which we will argue below does not seem to be the case.

[16]This development was also observed in our graphical analysis, which gives confidence that our estimation strategy captures well the trends of the outcome variables.

5.2. Is the increase in employment due to transitions into low quality jobs?

As previously seen, the reform seems to have fostered the employment among the disabled. However, this positive indicator of the success of the reform could be masking other factors, such as a worsening of the job quality of those who move to employment. That is, if disqualified DI claimants have rushed to employment as a response to the sudden income drop caused by the reform, then the quality of the jobs they find can be lower due to the reform. To get a better idea on the job quality, we follow Nekoer and Webei (2015) and assess the quality of employment by looking at the wages, the employment spell length and days in unemployment for those who first transitioned to employment after the reform. The two latter measures are a proxy for quality of employment by assessing its temporary or permanent nature. This is of relevance, as if employers are wary of hiring employees with a history of disability, they may be more likely to offer temporary employment contracts.

Table 6: Job quality of those transitioning to employment

	Job quality indicators
Wage	99.57**
	(49.20)
Calendar days worked	3.29***
	(0.36)
Days in UI	2.52***
	(0.15)

Note: This table shows the aggregated estimates of the triple difference-in-difference specified in equation (6)*. Significance stars indicate significance of the estimate at the 1 percent level ***, 5 percent level **, 10 percent level *. Standard errors are in parentheses, clustered at the individual level. The observations correspond to the number of individuals in the first period observed.*

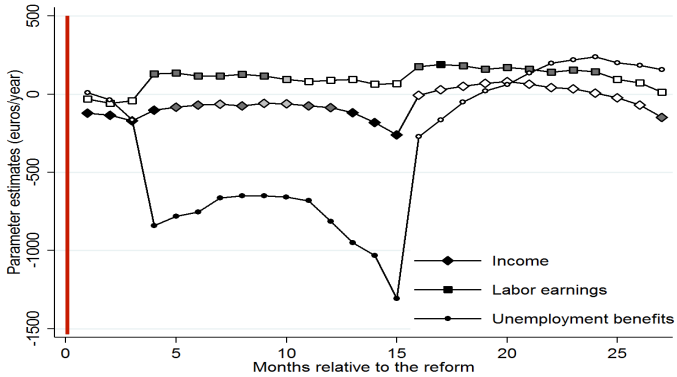
Table 6 shows the impact of the reform on employment quality indicators. The point estimate of the impact of the reform on wages is an increase of 99.6 euros a year. This increment is economically small, but provides evidence that treated claimants clearly do not experience employment with lower wages than their non-treated counterparts. Regarding the calendar days worked, it seems to be that treated individuals transitioning to employment work an average of 3.3 days more a year, and spend in unemployment 2.5 days more a year. Although statistically significant, this results are economically not really meaningful. We interpret this as the job quality not being different due to the reform. In this sense, we dismiss the hypothesis that the reform fostered the transitions to bad employment, but rather it fostered the transitions to the conventional jobs of people with disability histories.

5.3. What are the total income effects?

Overall, the analysis on the transitions to employment and non-employment states shows that non-eligible individuals tend to move to employment rather frequently. Also, we saw that those who do move to employment do not seem to be worse off in terms of employment quality than their non-treated counterparts. One could think that the administration did a good job at preventing a sudden income drop for the reassessed individuals, allowing them to take the time to find suitable jobs. In this section, we aim at shedding some light on this issue by investigating the total income effects. We do so by first looking at the yearly income of those eligible for reform, which we define as the sum of yearly labor earnings, yearly social programs benefits and self-employment earnings. Particularly, we perform specification (7) on income. Figure 4 shows the result from this estimation.

We note that during the first 15 months after the onset of the reform, the reform seems to have induced an income loss in the treatment group, ranging from 65 to 260 euros a year. However, from month 15 to month 25, the reform does not seem to have had any impact in the income of the treated individuals. To understand what are the drivers of this initial income loss, we also perform a flexible analysis of the reform on the labor earnings and the unemployment benefits received. Figure 4 shows the estimates for these two outcomes. Clearly, the drop in income in the initial years of the reform seems to be a consequence of the drop in unemployment benefits, which are 900 to 1200 euros a year lower. Contrasting this finding with the previous evaluation of the state transitions, it seems that although treated claimants transitioned more often to UI, the benefits received per person were on average lower. In that sense, one could argue that the leakage from DI to UI did not lead to an expansion of the costs of the UI program

Figure 4: Flexible estimates of the total income, labor earnings and unemployment benefits



Note: This figure shows the flexible estimates from specification (7)*, with clustered standard errors at the individual level. The horizontal axis, 0 corresponds to October 2004, the onset of the reassessment process. The estimates for each month after the onset of the reform are then plotted. The color of the markers indicate the significance level of the estimates. Black markers indicate a significance of the estimate at the 1% confidence level, dark gray at the 5% confidence level, light gray at the 10% confidence level, and white indicates non-significance.*

for the government, but did generate an income loss for the treated DI claimants. Note however, that we do not have information on the income received from the TRI program. If as specified by the UWV, the TRI ensures the pre-reassessment wage to all reassessed DI claimants, we should not observe an income loss at all. We can thus hypothesize that the UI program was less generous for the reassessed claimants, but only because of the preventive temporary income loss program put in place. Our results only allow us to say is that there was scope and need for such a temporary income program, at least in the short run. In the long run, instead, the income loss seems to have been mitigated. This seems to be driven by an increase in the labor earnings, which offsets any income effect of the reform. This result confirm the previous findings regarding the wages of the treated individuals transitioning to employment: the average wage earned is about 200 euros a month higher than in the absence of the reassessment reform, and converges to zero by the end of the observation period.

From our analysis regarding the income effects of the reform, we can conclude two things. First, there was an income drop in the first months following the onset of the reform, mainly caused by a drop in the unemployment benefits paid. Second, this income drop disappears around 15 months after the implementation of the reform, driven by an increase in the wages earned. However, because our data does not contain information on the payments from the TRI program, we do not know if the initial income drop estimated was actually realized, of if the TRI payments compensated for it. We can only conclude that the administration was right in putting in place such a program, as there is scope for income losses from the reform, albeit being relatively small.

5.4. What are the short-term health effects?

We investigate in this subsection the health status of the treated claimants. Although they do not seem to be doing worse in terms of income and the employment opportunities offered to them, if their health status is becoming worse due to the reform, we must consider how positively we should evaluate it. When it comes to public welfare, a fall in the population health status is very undesirable, as besides its moral implications it also comes associated with higher public health expenditures. To do so, we use hospital registries data from the National Medical Registries (Landelijke Medische Registratie (LMR)). The data is available from 2000 to 2005[17], and we report three outcomes: the first one is a binary variable for being hospitalized in a particular month, the second one are the nights of hospitalization, and the third one is the probability of an emergency hospitalization. Table 7 reports the results of estimating specification (6) on this set of hospitalization outcomes.

[17]Prior to 2006, the legislation obliged hospital to report all hospitalizations, which were registered in the LMR dataset. After that date, reporting was no longer mandatory, so a selection arises in terms of which hospitals decided to report their outcomes and those who didn't. For greater detail on this data, please see García-Gómez, van Kippersluis, O'Donnell, van Doorslaer (2013).

Table 7: Estimated hospitalization outcomes. Full sample and diagnosis group analysis.

	Full Sample
Outcome Variables	
Probability of hospitalization	0.23***
	(0.04)
Nights hospitalized	0.02**
	(0.01)
Probability of an emergency hospitalization	0.13***
	(0.03)
Observations	653,933

Note: This table shows the aggregated estimates of the triple difference-in-difference specified in equation (6)*. Significance stars indicate significance of the estimate at the 1 percent level ***, 5 percent level **, 10 percent level *. Standard errors are in parentheses, clustered at the individual level. The observations correspond to the number of individuals in the first period observed.*

The results show that there seems to be a positive impact of the reform on the probability of being hospitalized, of 0.23 percentage points. We do not believe this result should be given any economic interpretation, given its small magnitude. Similar conclusions arise from looking at the nights hospitalized and the probability of undergoing an emergency hospitalization: significant estimates, but of little economic interpretation. Overall, we conclude that there are no signs of changes in the hospitalization outcomes induced by the reform. Those results are far from conclusive, however: a much longer time horizon should be analyzed to find any effects, both from DI generosity to health, and from health outcomes to hospitalizations. Garcia-Gomez and Gielen (2014) find that there are negative impacts from reducing DI generosity on health outcomes, but only over 10 years after the onset of the reform. Our analysis only allows us to draw conclusions on the health outcomes one year after the reform, and we unsurprisingly do not find any negative impact from the reform.

5.5. Do the effects differ by initial health conditions?

In this subsection, we explore the heterogeneity across individuals with different health conditions. Recall that the medical criteria of the the new eligibility rules employed in the reform was left unchanged; Technically, the administration shouldn't have discriminated between health conditions. We then interpret the differences on DI exit rates across health conditions as the beliefs of the administration regarding the remaining work capacity associated to the different health conditions. The ratio between the probability of being in a particular state and the probability of leaving DI hinges on the different levels of error type II, moral hazard or labor market discrimination across diagnoses.

We estimate the baseline specification (6) and the dynamic effect specification (7) on six different diagnoses: general diagnosis, cardiovascular diseases, musculoskeletal disorders, mental disorders, neurological disorders, respiratory diseases and digestive. All chosen conditions represent a sizable fraction of the overall sample, allowing us to maintain the power of the analysis. General diagnosis, cardiovascular diseases, neurological, respiratory and digestive disorders are medically verifiable in their nature. Instead, musculoskeletal disorders, together with mental disorders, belong to the category of "subjective" health conditions. The point estimates of the static estimation can be found in table 8. We note that the point estimates differ significantly across diagnosis. Claimants suffering from subjective health conditions experienced the largest increases in the DI exit rate: it seems that the administration has considered this health conditions to be the most fitted for the labor market. Transitions to employment, however, do not seem to follow for those with mental disorders. Indeed, individuals with respiratory diseases seem to have transitioned to employment more assiduously. Regarding the transitions Indeed, these are often below 5% the magnitude of the other estimates, giving them very little economic interpretation, despite its statistical significance. We are wary of making conclusions on such small effects, and the reason will become clear as the placebo tests will be presented: the large sample size and small standard errors from our analysis often produce statistically significant results on almost negligible effects of the reform. In what follows, we will then focus on the transitions to employment, unemployment and other non-employment states.

We report in figure 5 the estimated effect on ratio between the transition probabilities to those states, both for the static estimation approach, and for the dynamic approach from specification 7. Figure **??** in the appendix shows the

Table 8: Estimates of the impact of the reform on the DI exit rate and transition probabilities by Diagnosis

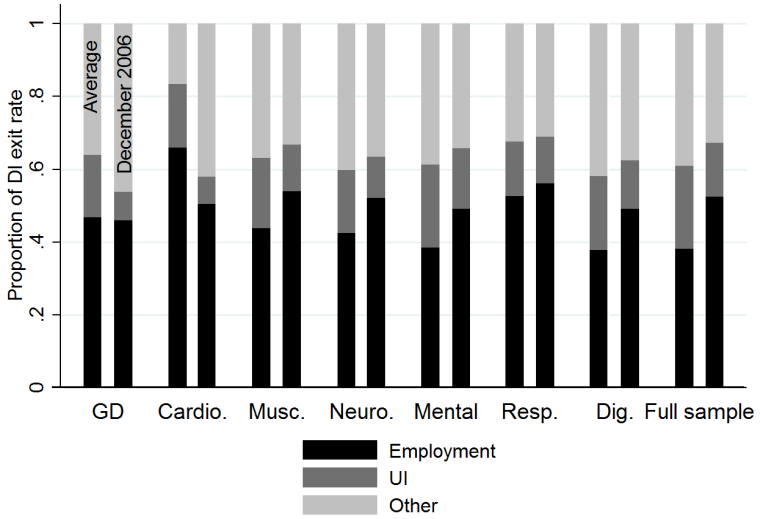
	General	Cardiovascular	Musculoskeletal	Mental	Neurological	Respiratory	Digestive
Probability of being							
on Disability	-3.19***	-1.48**	-5.57***	-5.38***	-3.95***	-4.81***	-4.95***
	(0.29)	(0.54)	(0.2)	(0.18)	(0.41)	(0.77)	(0.7)
Employment	1.14***	0.95**	2.46***	1.88***	1.48***	2.17***	1.47***
	(0.21)	(0.39)	(0.15)	(0.13)	(0.31)	(0.58)	(0.55)
Unemployment	0.51***	0.25*	1.08***	1.12***	0.60***	0.62**	0.79***
	(0.07)	(0.14)	(0.05)	(0.05)	(0.1)	(0.24)	(0.21)
Self-Employment	0.20***	0	0.21**	0.25***	0.03	0.07	0.33**
	(0.06)	(0.11)	(0.05)	(0.04)	(0.08)	(0.17)	(0.16)
New Disability spell	0.26*	-0.45	-0.25	0.25**	0.54***	0.20	0.23
	(0.15)	(0.29)	(0.16)	(0.1)	(0.2)	(0.33)	(0.27)
Other	1.08***	0.24**	2.07***	1.89***	1.14***	1.34***	1.63***
	(0.12)	(0.13)	(0.08)	(0.08)	(0.23)	(0.44)	(0.37)
Observations	87,751	19,720	194,849	239,622	35,030	10,231	14,305

Note: This table shows the aggregated estimates of the triple difference-in-difference specified in equation (6) *by diagnosis subgroups, with standard errors clustered at the individual level. The estimates have been multiplied by 100, so that they are readily interpretable as percentage points variations. Significance stars indicate significance of the estimate at the 1 percent level ***, 5 percent level **, 10 percent level *. Standard errors are in parentheses. The observations correspond to the number of individuals in the first period observed.*

dynamic effects from the latter. Within each diagnosis group, the bar on the left in figure 5 shows the decomposition of the DI exit rate estimated through the baseline specification (6). The bar on the right decomposes the DI exit rate in December 2006, obtained from the flexible specification estimates. For both bars, the effect on the DI exit rate is set to be the unity, and the proportions of the effects on the transition states are adjusted to add up to the total DI exit rate. Reporting the estimates with this figure allows us both to compare the average impact of the reform and its effect in the longer run, thus getting a snapshot on how the proportion of each transition probability evolved since the onset of the reform.

If we first look at the average impact of the reform on the different diagnosis groups, we note that claimants with mental ill-health experience among the lowest relative transitions to employment, together with individuals suffering from diseases of the digestive system experience. Both proportions seem in fact quite close to the average employment fraction in the full sample. Individuals suffering musculoskeletal disorders experience a higher relative transition to employment than the sample average, but well below those with cardiovascular and respiratory diseases. Regarding the transitions to UI and other non-employment states, claimants suffering from mental disorders experience the highest relative increase, again close to the one for the full population sample. How do those proportions compare to the long term estimated proportions? We note several main points. The first one is that individuals with mental disorders experience one of the largest increases in the fraction of employment over time. By the end 2006, disqualified claimants with mental disorders have increased their relative labor participation from 38% to 50%. This change is among the largest, but still after that of those with digestive disorders, and that of the average treated claimant. It seems that claimants suffering form mental diseases start with comparatively low transitions to employment, but converge to the level of the other diagnosis groups over time. Second, we note a decrease in the long run ratio of employment to the DI exit rate for those suffering with cardiovascular diseases. This result should be taken with care, given the relatively small face value estimates obtained for claimants with this health condition. Table 8 shows that the estimates for this subgroup are 50% or more smaller than the ones for other health groups. Given its small size, we are careful when considering those results. Regarding the relative transitions to UI, we note the particularly high average effect of the reform for those suffering from mental disorders. When it comes to the long term effect, the impact of the reform on the relative probability to transition to UI seems to have diminished and converged within diagnosis groups. Thus, individuals suffering from mental disorders seem to have retained a higher eligibility to UI benefits than those from other diagnosis groups, suggesting that their high employment level at the onset of their disability. This seems to contradict previous evidence on the labor market behavior of individuals with mental diseases, who are believed to have much lower employment rates (OECD, 2014).

Figure 5: Estimated fraction of the main transition probabilities over the total DI exit rate by diagnosis



Note: This figure shows the estimated ratio between the transition probabilities to employment, UI and residual non-employment states and the DI exit rates for the baseline specification (6) *(left bar), and in December 2006, obtained from the flexible specification estimates* (7) *(right bar). The effect on the DI exit rate is set to be the unity, and the proportions of the effects on the transition states are adjusted to add up to the total DI exit rate.*

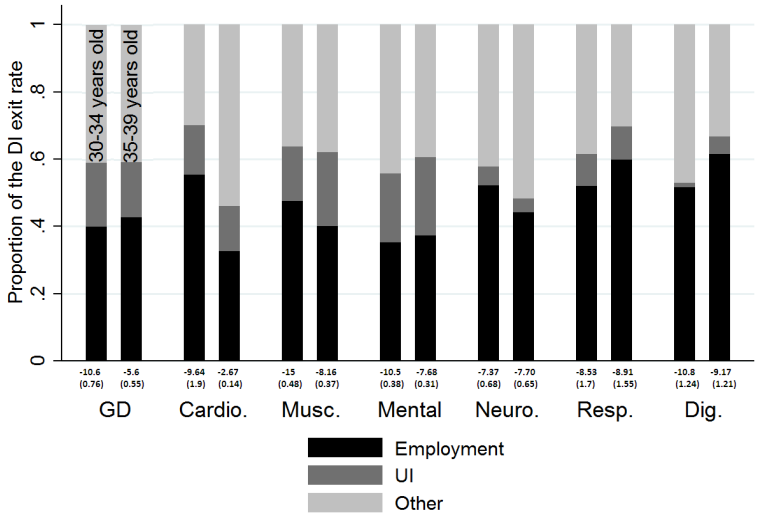
Our previous findings seem to highlight significant differences across health conditions. However, one has to bear in mind that these differences could be partly caused by the underlying differences across diagnosis groups in terms of age. In fact, if disproportionately younger people suffers from a particular health condition the cohort implementation of the reform would lead us to find stronger effects for that particular condition. Table 10 in the appendix shows the results from exploring the heterogeneity of the results by age groups within each diagnosis group.[18] We present in figure 6 the proportions resulting from the estimation of the static trend adjusted diff-in-diff approach, for the age groups 30-34 and 35-39 in July 2004. The estimated point estimates for the DI exit rate can be found in figure 6, under the x-axis. We first note that when looking at the younger cohort, the face value of the effect on the DI exit rate almost triples. We suspect that the increased frequency of reassessments and higher assessed suitability for the labor market for younger claimants play an important role when examining the results. We also note that within a same age group, the variation across health conditions in the estimated effects decreases. So, conditional to belonging to a certain age group, the administration beliefs regarding the suitability for the labor market of the different health conditions converge.

We note that there is some variation in the relative transitions within a groups of diagnosis, between groups of age. This variation is however small, and amplified by the re-scaling of the proportion rates, and the increased non-statistical significance of the coefficients. If we focus on the diagnosis groups with the larger number of observations and more sizable effects of the reform, i.e leaving aside individuals with cardiovascular diseases, we see that the proportions are rather stable across age groups. This seems to point to the fact that individuals of those two age groups responded similarly to the reform. Also, the proportions obtained accounting for heterogeneous age effects are similar to the ones reported in figure 5. We can more confidently attribute the different effects between health conditions to the health conditions themselves, rather than to underlying age composition effects.

In sum, we conclude that the reform had a stronger impact in the probability of leaving DI for those who suffered from musculoskeletal disorders. Individuals with mental disorders do not seem to have experienced stronger increases in the DI exit rate, when we account for age composition. The administration does not seem to have disqualified

[18] We focus only on the youngest eligible groups. During the span of the analysis, fewer reassessments took place among older individuals, resulting in estimates of the impact of the reform lacking of economic meaning. The analysis for the subgroups of older individuals is available under request.

Figure 6: Estimated fraction of the main transition probabilities over the total DI exit rate by diagnosis and age group



Note: This figure shows the estimated ratio between the transition probabilities to employment, UI and residual non-employment states and the DI exit rates for the baseline specification (6) *per diagnosis and age groups. The effect on the DI exit rate is set to be the unity, and the proportions of the effects on the transition states are adjusted to add up to the total DI exit rate. The estimated point estimates for the DI exit rate can be found under the x-axis, with standard errors in parentheses.*

particularly those with mental disorders. This outcome seems to go in line with that of the other reforms of the Dutch DI scheme in the early 2000's: evidence shows that these reforms have been much more successful at curbing the entry to DI from all other kind of diagnoses than mental disorders (UWV, 2012). The extent to which we consider this a negative outcome depends on the relative transitions to employment, UI and non-employment states. We have seen that ineligible DI claimants suffering from mental disorders present below average transitions to employment. In light of such low retained employment capacity, we wonder whether a stricter treatment for the ones with mental ill-health would have been welfare improving. On the other hand, we also saw that the transition to employment to years after the reform shows a much more positive outlook on the transition to employment for those with mental health: the relative employment response converges to that of the other health conditions. Thus, claiming that those with mental disorders have a lower retained employment capacity may be short sighted: rather, they seem to take more time to transition to employment. This can be in fact linked to a higher retained eligibility to UI. If disqualified DI claimants claim their entitled UI benefits, and transition to employment only after exhaustion of their rights, the fact that those with mental disorders had a higher entitlement would postpone their transition to employment. We cannot conclude whether this is due to a high disutility from working, or to using the time in UI to search for a suited job.[19]

5.6. Do the effects differ by duration in DI?

The duration of disability spells may have a direct impact on the claimants' ability to retake employment, by causing skill deterioration, social exclusion, depression, or lack of self-esteem (Vingard et al. (2004), Bryngelson et al. (2009), Svensson et al. (2010)). This could turn able DI applicants into unable to work after having spent significant time receiving the benefits. We explore this hypothesis by performing the following estimation:

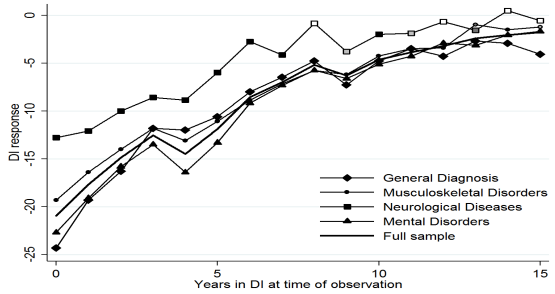
$$y^{state}_{igptd} = \delta D_g T_t + \theta_d D_g T_t spell_d + \sum_{p=-20}^{26} m_p + D_g \sum_{p=-20}^{26} m_p + \mu_i + \gamma_t + \epsilon_{igpt} \tag{8}$$

[19] Although an analysis of the impact of the reform on the wage for those with mental disorders produces an estimates of 168 euros more a year (s.e 44), 70% that above the average effect.

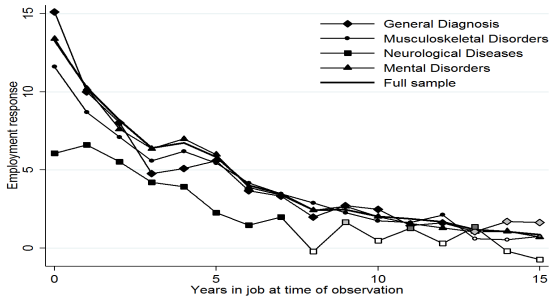
Where (8) is an extension of specification (6), in which a set of dummies for the different DI spell lengths $spell_d$ is interacted with the treatment effect $D_g T_t$. To calculate the duration on DI of each individual, we have measured the time spent on DI at the moment of observation (January 2000 for the individuals in the comparison period, and January 2003 for the individuals in the reference period) and rounded it at the yearly unit. We then include 16 duration dummies, ranging from zero years to 15 years on DI. We stop the analysis at 15 years on DI, where more than 75% of our sample is included. We allow the parameter of the interaction between the duration and the treatment effect θ_d to vary per duration on DI d. Figure 7 plots the estimated treatment coefficients δ and θ_d as a function of duration on DI, for some of the most relevant diagnosis groups and the full sample.

Figure 7: Analysis of the DI exit rate and main transition states per years on DI

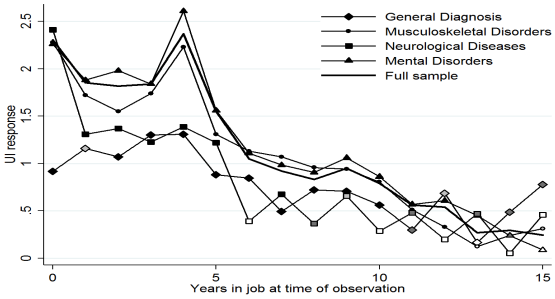
Probability of being on DI



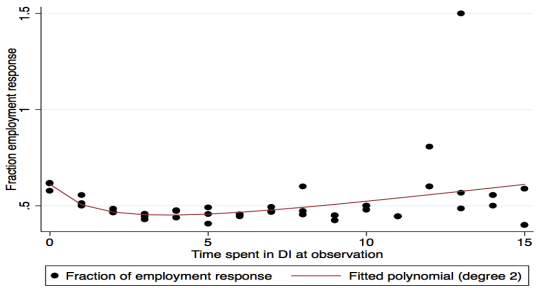
Probability of transitioning to employment



Probability of transitioning to UI



Employment response relative to DI exit rate response



Note: This figure shows treatment effect estimates from specification (8) *for different diagnosis groups, as a function of the time spent on DI. The color of the markers indicates the significance level of the estimates. Black markers indicate a significance of the estimate at the 1% confidence level, dark gray at the 5% confidence level, light gray at the 10% confidence level, and white indicates non-significance. The parameter estimates have been multiplied by 100, so that they are readily interpretable as percentage point variations. The figure in the lower right corner plots the point estimates of the employment response and DI response for all disability causes from figure 7 as a fraction of each other. We have fitted a second degree polynomial function to the scatterplot.*

The impact on the exit rate from DI decreases with the time spent on DI, and seems to converge to a null impact on the probability of being on DI. This can be due to the fact that individuals having spent longer periods in DI are sicker, thus the reassessments had a smaller impact on the probability of leaving DI. Alternatively, it may also be capturing the fact that during the period under analysis, little reassessments took place in the group of older individuals, who are most likely those to have stayed the longest on DI. The impact of the reform on the transition to employment is decreasing over time spent on DI. To investigate this, we plot the evolution of the employment response as a fraction of the total DI rate response, to which we fit a second degree polynomial line. We see that overall, the negative relationship between the employment response and the time on DI becomes U-shaped when measuring it relative to the DI exit rate. One has to be careful with this, as it is strongly sensitive to the degree of the polynomial function chosen to fit the data: we do not want to give any additional interpretation to the shape obtained from the fitted prediction. We do want to note, however, the similarity of the employment response as a fraction of the DI exit rate across all diagnosis groups. Also, that the decreasing trend of the employment response with the time in DI observed

in figure 7 is mainly driven by the impact of the DI exit rate, rather than the actual time spent on DI. Regarding the probability to transitioning to UI, we find a much more erratic pattern, but the evolution is similar to that of the transition to employment. We also assume that this pattern is strongly driven by the effect of the reform on the exit rate from DI, rather than the actual time spent in DI.

5.7. Placebo Tests

To assess the plausibility that the strategy followed identifies the impact of the reform, we perform several placebo tests. First, we perform the specification defined in (6) by using age groups older than 50 years old, hence not eligible for the reform. Particularly, we define these claimants aged in between 50 to 53 in July 2004 as our treatment group. As in the specified trend-adjusted difference-in-difference, we select those in between 50 to 53 in July 2001 to be the control group 2, and those in between 54 to 55 in July 2004 to be the control group 1.[20] Table 9 shows the results of this test. We can see that we do not find significant results for any estimate other than for the transition probability to UI. This result may not be surprising, given the previous graphical evidence. The treated cohort seems to be undergoing different trends in the transition probability to UI, invalidating the no cohort effects assumption. The reported estimates on the transition probability to UI should then be taken with care.

Table 9: Placebo tests on cohorts of older individuals

	Young: 50-53. Old: 54-55.
Outcomes	
DI	0.14
	(0.14)
Employment	-0.08
	(0.11)
UI	-0.08*
	(0.04)
Others	0.04
	(0.08)
Observations	249,837

A second placebo test we have performed to test the empirical strategy proposed in (6), is to set placebo reforms in each month of from March 2003 to July 2004, and analyze the response of the outcomes of interest. The graphical representation of the point estimates can be seen in figure 8. In the horizontal axis, we plot the months in which the reform has been set, where 0 corresponds to January 2003 (January 2000) for the reference (comparison) period. The first placebo reform is set to be in March, and continuously every month until July 2004. This graph shows the 95% confidence intervals of the estimates. We note several points. First, the estimates are generally statistically significant. This is worrying, as it would suggest that our strategy captures an impact from our placebo reforms, thus invalidating the approach. On the other hand, the estimated effects range from -0.5 to 0.5 percentage point increases, which is are very small effects. We are able to identify such small effects due to the large sample size we have, which produces extremely small standard errors. By looking at the confidence intervals of each estimate, we can evaluate the large precision of our estimates. We draw a lesson from this evidence, and it is that significance does not have to be confused with economic relevance, particularly in our analysis. It is a call for the discretion of the researcher and the reader to consider which effects should be considered as economically meaningful, rather than to blindly rely on statistical significance. By focusing on the sign of this placebo estimates, we also want to note that they all point to an overestimation of the impact of the reform both on the DI exit rate, and the transition probabilities to employment and non-employment. However, this overestimation is likely to be small: our argument is supported by the proximity of our estimates with most of the previous estimates of the literature.

Finally, a last placebo test will consist in evaluating the flexible empirical strategy, as defined in specification (7). To do so, we set a placebo reform at month -10 prior to the reform and analyze the dynamics of the impact of the placebo reform up to month -1 relative to the reform. We set the placebo reform at this point to have the same

[20] We do not select individuals older than 55 years old to follow in the reference period (January 2003-December 2006), since their labor market decisions could be affected by the reform in the Dutch pension system in 2004. If this is so, it would clearly invalidate the no-cohort effects assumption.

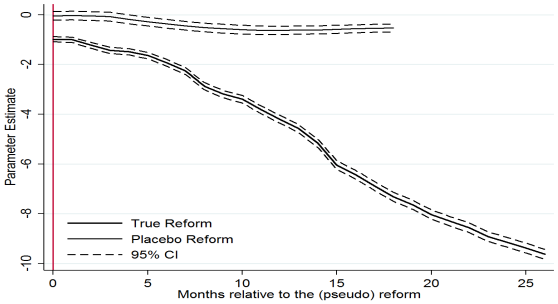
Figure 8: Placebo reforms on the different outcomes of interest



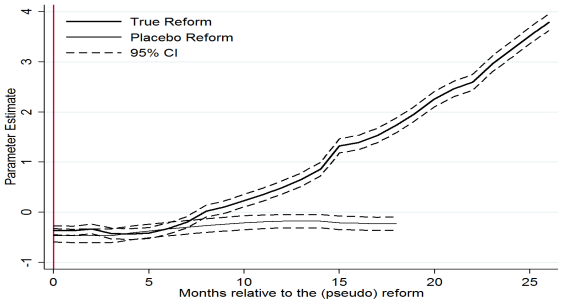
proportion of pre- and post-reform months than with the true reform. Figure **??** shows the results of this estimation for the probability of being on DI, transitioning to employment and UI. This figure shows both the estimated impact of the true reform, also reported in figure 3, and the estimated impact of the placebo reform. The horizontal axis corresponds to the months relative to the (placebo) reform, which is set at the origin. For all outcomes analyzed, we can see that the estimates of the placebo reform are extremely close to zero as compared to the estimates of the true reform, and follow a straight horizontal line. Again, some of the point estimates are statistically significant, but just by comparison with the actual period of the reform, it is difficult to argue that the placebo reform had some impact on the probabilities analyzed. This adds up to the previous discussion, which considers not only economic significance as the ultimate benchmark for assessing the impact of the reform.

Figure 9: Placebo test flexible estimates

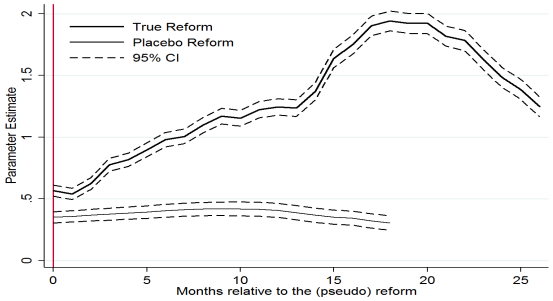
Probability to be on DI



Transition probability to employment



Transition probability to UI



6. Conclusion

This paper has brought evidence on the employment response of DI beneficiaries faced with a stricter criteria to qualify for DI. The study of this particular reform is interesting, as it provides a set up to test the expectations of the administration regarding the residual capacity of employment of DI claimants with the actual employment response, as well as with the transitions to non-employment states. The findings suggest that the reform was a success, at least in the short run. In terms of fostering employment, we find that two years after the onset of the reassessments, 50% of the DI exit rate was reallocated to employment. The reform did not represent an income loss for those eligible for reassessment, neither a transition to worse quality employment. We find no major evidence of discrimination in the labor market due to the previous reforms. Although there are some movements towards other social assistance programs, we do not observe an increase in the benefit receipt from those.

Claimants suffering from mental disorders overall experience a quite close to average transition to employment, but do take longer in becoming employed. We are not able to identify a negative impact from disability benefit receipt that worsens with the time a claimant has spent on DI. This result holds for all types of diagnoses.

References

Autor, David H. and Duggan Mark G. (2003). The Rise in the Disability Rolls and the Decline in Unemployment Quarterly Journal of Economics, 118(1), February 2003, 157 - 206.

Autor, David H. and Duggan Mark G.(2006). "The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding." Journal of Economic Perspectives, 20(3): 71-96.

Bell, B., Blundell, R., & Van Reenen, J. (1999). Getting the unemployed back to work: the role of targeted wage subsidies. International Tax and Public Finance, 6(3), 339-360.

Borghans, L., Gielen A., & Luttmer E. (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". (No. w2816). National Bureau of Economic Research.

Burkhauser, R. V., and Daly, M. C. (2011). Testing educational tools to demonstrate returns to work for children aging out of the SSI-disabled children program.

Bryngelson, A. (2009). "Long-term sickness absence and social exclusion." Scand J Public Health 2009, 37(8), 839-845.

Campolieti, M. and Riddell, C., (2011). "Disability policy and the labor market: evidence from a natural experiment in Canada, 1998-2006", Journal of Public Economics 96, p.306-316.

Charles, K. K. (2003). The longitudinal structure of earnings losses among work-limited disabled workers. Journal of Human Resources, 38(3), 618-646.

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

French, E. (2005). The effects of health, wealth, and wages on labour supply and retirement behaviour. The Review of Economic Studies, 72(2), 395-427.

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

Duggan, M., & Imberman, S. A. (2009). Why are the disability rolls skyrocketing? The contribution of population characteristics, economic conditions, and program generosity. In Health at older ages: The causes and consequences of declining disability among the elderly (pp. 337-379). University of Chicago Press.

Duggan, M. (2015). Senate testimony. `http://www.budget.senate.gov/democratic/public/_cache/files/3ceeaf1b-9427-48c2-a51b-e422c36eb802/mark-duggan-testimony.pdf`

Garcia-Gomez, P. & Gielen, A. C. (2014). Health Effects of Containing Moral Hazard: Evidence from Disability Insurance Reform, Tinbergen Institute Discussion Papers 14-102

Hassink W., Koning P., Zwinkels W. (2015). Employers opting out of public disability insurance: selection or incentive effects?. IZA discussion paper, IZA DP No. 9181.

Karlstrom A., Marten P., Svensson I. (2008). The employment effect of stricter rules for eligibility for DI: evidence from a natural experiment in Sweden. Journal of Public Economics, April 2008, pages 2071-2082.

Koning, Pierre, 2009, Experience Rating and the Inflow into Disability Insurance, De Economist, 157(3), 315-335.

Koning, Pierre, 2012, Anticipated and Unanticipated Incentives of Disability Insurance Experience Rating: The Case of The Netherlands, Policy and Practice in Health and Safety, 10(2), 63-76..

Koning, P., & Lindeboom, M. (2015). The Rise and Fall of Disability Insurance Enrollment in the Netherlands. The Journal of Economic Perspectives, 29(2), 151-172.

Kostol, A. R., & Mogstad, M. (2013). How financial incentives induce disability insurance recipients to return to work (No. w19016). National Bureau of Economic Research.

Lahiri, K., Song, J., & Wixon, B. (2008). A model of Social Security Disability Insurance using matched SIPP/Administrative data. Journal of Econometrics, 145(1), 4-20.

Li, J., Hurley, J., DeCicca, P. and Buckley, G. (2014), Physician Response to Pay-for-performance: Evidence from a Natural Experiment. Health Econ., 23: 962–978. doi: 10.1002/hec.2971

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

Moore, T. (2015), The employment effects of terminating disability benefits, Journal of Public Economics, Volume 124, April 2015, Pages 30-43

Mora Villarrubia, R., & Reggio, I. (2012). Treatment effect identification using alternative parallel assumptions. Universidad Carlos III, Departamento de Economia.

O'Donnell, O., van Doorslaer, E., & Van Ourti, T. (2013). Health and Inequality, Netspar Discussion Paper No. 10/2013-060.
Staubli S. (2011). The impact of stricter criteria for disability insurance on labor force participation. Journal of Public Economics, Volume 95, May 2011, 1233-1235.

OECD. (2010). Sickness, Disability and Work: Breaking the Barriers: A Synthesis of Findings across OECD Countries, OECD Publishing, Paris.

OECD. (2012), Sick on the Job?: Myths and Realities about Mental Health and Work, Mental Health and Work, OECD Publishing, Paris.

Svensson T, Müssener U, Alexanderson K. (2010). "Sickness absence, social relations, and self-esteem: a qualitative study of the importance of relationships with family, workmates, and friends among persons initially long-term sickness absent due to back diagnoses." Work, 37(2):187- 197

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

UWV (Uitvoeringsinsituut Werknemersverzekeringen), 2008. "Staatsblad 2007", Amsterdam, The Netherlands.

UWV (Uitvoeringsinsituut Werknemersverzekeringen), 2006. "UWV, 4e kwartaal 2005", Amsterdam, The Netherlands.

UWV (Uitvoeringsinsituut Werknemersverzekeringen), 2009. ”Rapportage afronding eenmalige herbeoordelingsoperatie”, Amsterdam, The Netherlands.

Vingard E, Alexanderson K, Norlund A. (2009). ”Consequences of being on sick leave.” Scand J Public Health, 32 (63 suppl), 207-215.

Von Hippel, Paul T. (2005). ”Mean, Median, and Skew: Correcting a Textbook Rule”. Journal of Statistics Education 13

Von Wachter, T., Song, J., & Manchester, J. (2011). Trends in employment and earnings of allowed and rejected applicants to the social security disability insurance program. The American Economic Review, 3308-3329.

Appendix

Categories under General Diagnosis:

- Fever
- Pain
- Malaise and fatigue
- Other symptoms, no elsewhere classified
- Kaposi's sarcoma
- Other neoplasms
- Chromosome Abnormalities appointed in number and shape
- Congenital abnormality, not elsewhere classified
- Tuberculosis
- Status HN-positive
- Disease by HIV
- Disease Pfeiffer
- Tetanus
- Brucellosis
- Other zoonoses
- Late effects of certain infectious diseases
- Other specified infectious diseases
- Freezing
- Sarcoidosis [Disease Besnier Boeck]
- Other diseases
- Systemic disease of the connective tissue
- Metabolism of lipoproteins
- Other metabolic disorders
- Nutritional Deficiency
- Obesity [adiposity / obesity]
- Other forms of overfeeding
- Acute poisoning
- Chronic poisoning
- Complication of surgical and medical treatment
- Toxic inhalation fever
- Radiation Sickness
- Because of heat and light
- Effect of reduced temperature
- Caisson Disease
- Other effects of air pressure and water pressure
- Asphyxia
- Allergic reaction
- Other undesirable consequences of external causes
- Sickness unspecified

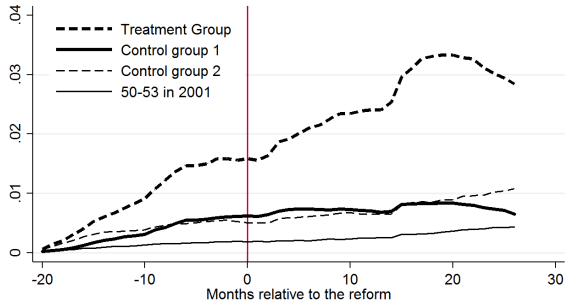
Table 10: Estimates of the impact of the reform on the DI exit rate and transition probabilities by Diagnosis and age

	30-34 years old							35-39 years old						
	GD	Cardio.	Musc.	Mental	Neuro.	Resp.	Digest.	GD	Cardio.	Musc.	Mental	Neuro.	Resp.	Digest.
Probability of being in														
Disability	-10.60***	-9.64***	-15.00***	-10.50***	-7.37***	-8.53***	-10.80***	-5.60***	-2.67***	-8.16***	-7.68***	-7.70***	-8.91***	-9.17***
	(0.76)	(1.9)	(0.48)	(0.38)	(0.68)	(1.7)	(1.24)	(0.55)	(0.14)	(0.37)	(0.31)	(0.65)	(1.55)	(1.21)
Employment	3.35***	5.49***	6.98***	3.44***	4.24***	4.01***	5.84***	2.05***	1.13	3.31***	2.65***	3.70***	3.87***	3.46***
	(0.58)	(1.40)	(0.39)	(0.28)	(0.58)	(1.40)	(1.09)	(0.41)	(0.99)	(0.29)	(0.22)	(0.54)	(1.34)	(1.05)
Unemployment	1.63***	1.47*	2.39***	2.00***	0.46*	0.74	0.15	0.80***	0.47	1.82***	1.66***	0.34	0.65	0.30
	(0.24)	(0.70)	(0.17)	(0.13)	(0.28)	(0.70)	(0.53)	(0.15)	(0.42)	(0.10)	(0.10)	(0.27)	(0.73)	(0.54)
Other	3.44***	2.97***	5.32***	4.33***	3.44***	2.97***	5.32***	1.96***	1.87**	3.14***	2.81***	4.33***	1.96***	1.87**
	(0.42)	(0.90)	(0.25)	(0.21)	(0.42)	(0.90)	(0.25)	(0.27)	(0.63)	(0.17)	(0.15)	(0.21)	(0.27)	(0.63)
Observations	33,825	9,703	77,771	87,884	6,838	1,548	2,651	38,585	10,427	87,183	99,496	7,518	1,754	2,925

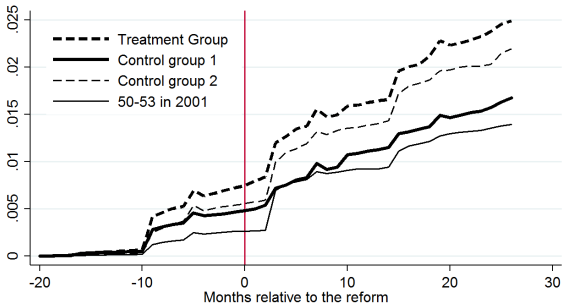
Note: This table shows the aggregated estimates of the triple difference-in-difference specified in equation (6) *by diagnosis and age subgroups, with standard errors clustered at the individual level. The estimates have been multiplied by 100, so that they are readily interpretable as percentage points variations. Significance stars indicate significance of the estimate at the 1 percent level ***, 5 percent level **, 10 percent level *. Standard errors are in parentheses. The observations correspond to the number of individuals in the first period observed.*

Figure 10: Trends in the transition probabilities to non-employment and self-employment

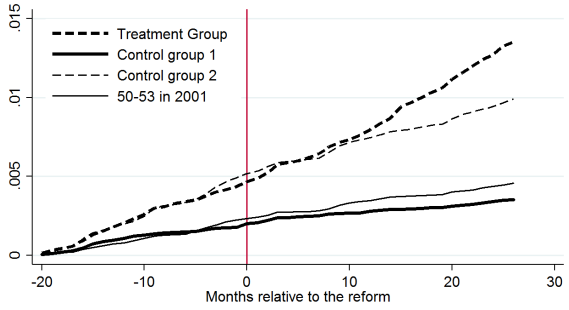
Transition probability to the UI program



Transition probability to a new DI spell



Transition probability to self-employment



Transition probability to any other state

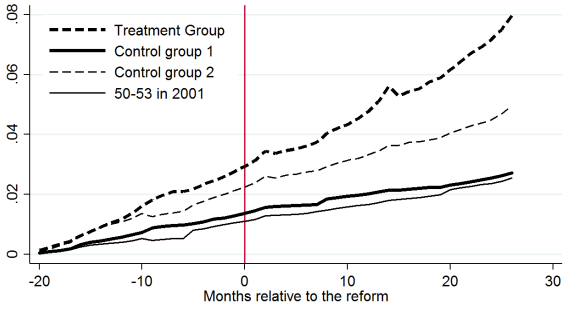
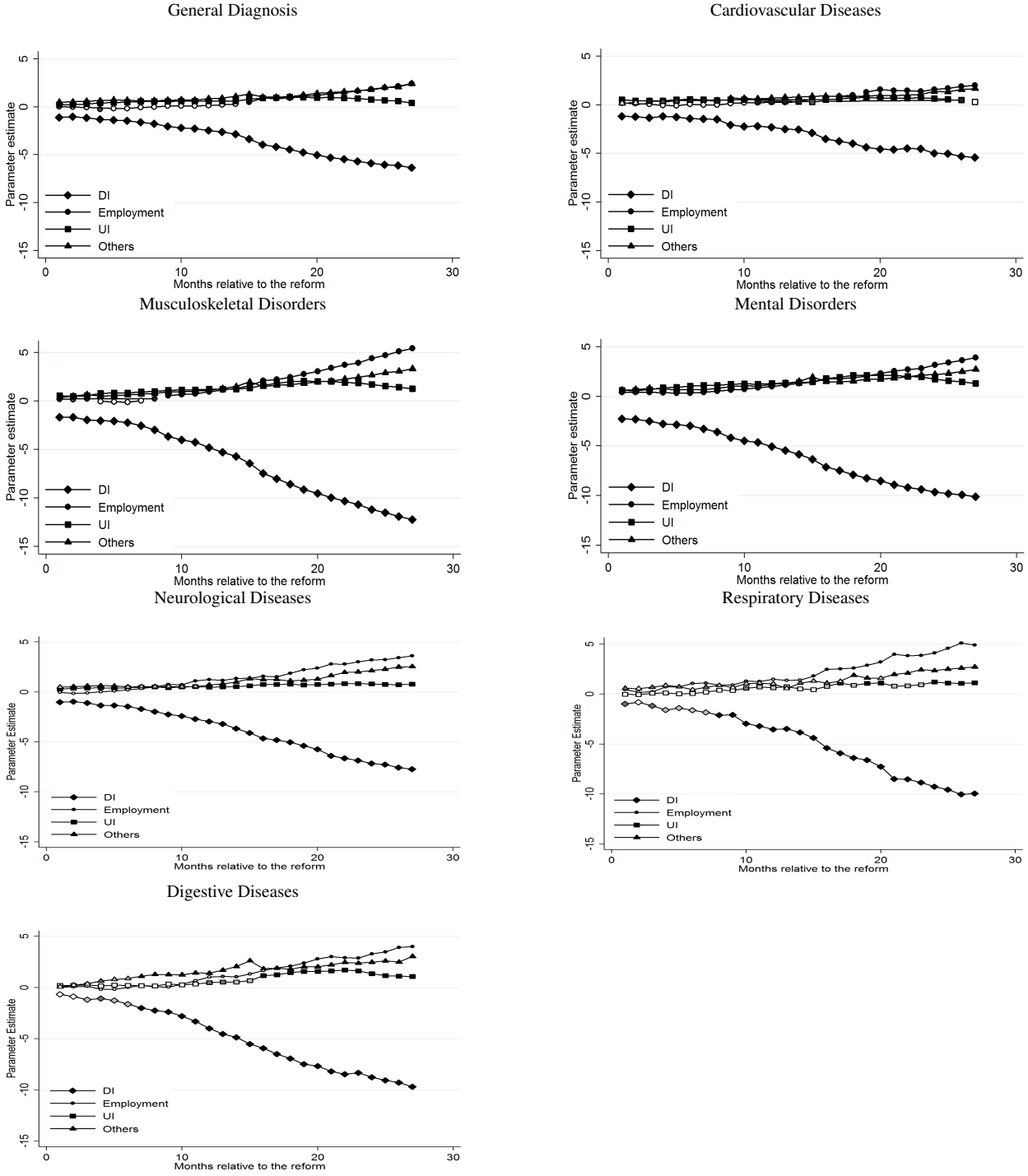


Figure 11: Flexible estimates of the impact of the reform on the DI exit rate and main transition probabilities by disability cause



Note: This figure shows the flexible estimates from specification (7) *for different diagnosis groups. The horizontal axis, 0 corresponds to October 2004, the onset of the reassessment process. The estimates for each month after the onset of the reform are then plotted. The color of the markers indicates the significance level of the estimates. Black markers indicate a significance of the estimate at the 1% confidence level, dark gray at the 5% confidence level, light gray at the 10% confidence level, and white indicates non-significance. The parameter estimates have been multiplied by 100, so that they are readily interpretable as percentage point variations.*