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

Candidate: Silvia Garcia Mandico

Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, Netherlands

Supervisor: Prof. Dr. Owen O'Donnell

Erasmus School of Economics (Applied Economics), Erasmus University Rotterdam, Burgemeester Oudlaan 50, PO Box 1738 3000 DR Rotterdam, The Netherlands

Supervisor: Dr. Pilar Garcia-Gomez

Erasmus School of Economics (Applied Economics), Erasmus University Rotterdam, Burgemeester Oudlaan 50, PO Box 1738 3000 DR Rotterdam, The Netherlands

Abstract

This thesis exploits a change in the stringency of eligibility of the Dutch DI scheme, which entailed the reassessment of a large fraction of DI claimants. I investigate the employment responses among those eligible for reassessment. I find that the employment response was of half the magnitude of that of the DI exit rate increase. This finding is consistent across groups of age and health conditions. There is evidence of strong market frictions in the short run, but one cannot dismiss the presence of moral hazard or type II errors in the award of DI benefits.

[☆]This document is written as MPhil Thesis for the Graduate Program in Economics of the Tinbergen Institute, academic year 2014-2015

August 24, 2015

Contents

1. Introduction

In July 2014, the US Social Security (OASDI) projected that the Disability Insurance Trust Fund would be depleted by 2016. In a Testimony before the Senate Budget Committee, Mark Duggan (2015) has attributed this predicament, in part, to the liberalization of the medical eligibility criteria for Disability Insurance (DI) in 1984 (as shown in in Duggan and Imberman, 2009). That led to higher application rates from individuals with "subjective" health conditions, such as back pain or depression. If individuals suffering from these conditions have a higher residual working capacity[1], an increasing fraction of able-to-work individuals are becoming part of the stock of DI claimants.

Since these disorders are associated to relatively low mortality rates, an increasing share of claimants suffering from these conditions raises the average duration of DI dependence (Autor and Duggan, 2006). In addition, because they are less age-dependent than heart diseases or cancer, to name a few, DI beneficiaries are becoming younger (Watcher et al. 2011).

In short, the US DI scheme has attracted a younger, healthier and poorer population. Similar developments led to major reforms of the Dutch DI scheme in the early 2000s. Most of these reforms aimed at reducing the incentives of entry and are considered to be game changers in curbing the inflows to DI (van Sonsbeck and Gradus (2012), Koning and Lindeboom (2015)). The focus of this thesis is a reform that has received far less attention. It involved the re-evaluation of a large fraction of DI recipients' cases under a stricter eligibility criteria. This led to a reassessment of over 345.000 beneficiaries under 45 years of age in 2004[2]. The new eligibility rules were more flexible in considering which potential jobs recipients could still perform despite their disability. Medical criteria was left unchanged. Over 20 percent of the reassessed were considered ineligible for DI, and 12 percent of those who remained in DI had their benefits reduced.

In this thesis, I explore the impact of this reform on the individuals eligible for reassessment. Particularly, I focus on the extent to which the reform fostered re-employment. Although the reform aimed directly at increasing labor market participation, one should not take as given that recipients loosing (part of) their DI benefit will compensate the income loss by working more. A reform that tightens the eligibility criteria to correct type I errors in DI awards may also increase the frequency of type II errors (Parsons, 1991). Therefore, employment may not increase one-to-one with DI roll decreases if the reform is disqualifying truly disabled applicants from DI.

Additionally, not all of the reduced type I errors will move back to work. Existing labor market frictions and discrimination against disabled individuals (O'Donnell et al. 2013) can cause employment to increase less rapidly than DI claims falls.

Also, rejected DI claimants with very low taste for work may prefer to move to other social assistance programs than to employment. Borghans, Gielen and Luttmer (2014) find that in the Netherlands, claimants exposed to a less generous DI benefit scheme have a 5.1 percentage points higher probability of being employed, and a probability of participation in other social assistance programs 7.8 percentage points higher.

In this thesis, I use administrative data maintained by Statistics Netherlands, containing the universe of disability claimants. The extensive and detailed nature of the data enables the identification of movements in and out of DI, and from and to employment. This allows me to explore the reform's effect on DI dependence and employment. The literature has often taken the approach to exploit discontinuities generated by policy reforms to recover the exogeneity of the reform on the marginally treated individuals (Borghans et al. (2014), Kostol et al. (2013)). Although this reform initially provided such a discontinuity, given that only these born after 1954 had to be part of the reassessment, the effective implementation of the reform invalidates the use of regression discontinuity techniques. Instead, I follow a trend adjusted triple differences-in-differences approach in the spirit of Bell et al. (1999). I define as my treatment group the claimants whose age in July 2004 qualified them for reassessment, and I follow them from 2003 to 2006. The control group are claimants who did not qualify for reassessment due to being 50 years of age or older in July 2004. To adjust for differential trends by age that may be expected to occur over a four-year period in the absence of the reform, I use the information on claimants of the same age as the treatment and control groups from a four-year period prior to the reform.

[1] Residual working capacity is defined as the remaining capacity to engage in a gainful activity despite suffering from an illness.

[2] Initially, the reform was meant to reassess all individuals below 50 years of age in 2004, but a law amendment reduced it to 45. This will be described in detail below.

I find that the reform induced an increase in the probability of employment of 3.7 percentage points. Claimants who were working and in DI at the time of the reform experience an increase in the probability of employment of 5.4 percentage points, whereas these only in DI experience an increase in 2.6 percentage points. Because the empirical strategy employed identifies the average impact of the reform on the population of DI claimants, I can explore the impact on a wide range of subpopulations. I investigate the employment responses of being exposed to a stricter DI eligibility criteria on different age groups. I find that the extent to which the employment response accompanied the rise in the exit rate from DI remains constant over groups of age. These results capture the short term employment responses to the reform, given that I only analyze up to two years after the onset of the reassessments. This is potentially exacerbating the gap between employment responses and the rise in DI outflow, if there were labor market frictions causing employment rates to rise sluggishly. The evolution of employment responses shows that this gap narrows over time, but still remains substantial even when the full impact of the reform has been reached. These results indicate that a reform specifically targeting the assessment of the employment possibilities of DI claimants may not result in one-to-one increases in employment and DI outflow. I interpret this finding as reflecting unwillingness to work among the treated claimants, as well as potential type II errors in the awarding of DI benefits.

The reform under analysis provides a particularly interesting set up to evaluate residual employment capacity by diagnosis. This is so, because the new reassessment criteria left the medical aspect unchanged, while reducing the amount of job requisites the claimant had to meet to be considered able to work. This implies that no medical conditions were specifically targeted for reassessment: differential DI exit rates increases across diagnoses should be interpreted as the administrations' beliefs on residual employment capacity differentials across health conditions. This provides a framework to investigate the difference between the administration expectations and the actual employment response. I find that the ratios between the boost in employment and the rise in DI outflow is the smallest for those claimants suffering from mental illnesses. Interestingly, this ratio is the largest for cardiovascular disorders, well above that of musculoskeletal disorders and mental diseases. The reform seems to have overestimated the employment capacity of these with musculoskeletal disorders and mental diseases, either by underestimating the labor market frictions they would face, the work discomfort of the condition, or how much they value work.

Despite being specific of the context of this reform, my findings add to the literature on the re-employment effects of disability insurance, and to the heterogeneous employment responses by health conditions. 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.

Watcher et al. (2011) find that, for the US, rejected DI applicants with musculoskeletal and mental disorders have the highest subsequent employment rates. Although these results are extremely informative, they cannot be interpreted causally, since rejected DI applicants can differ in unobservables from these accepted for DI (Lahiri, Song, and Wixon, 2008).

Maestas et al. (2013) exploit the variation arising from the different allowance propensities of DI case examiners to instrument being accepted on the US DI. They find that the employment rate of new beneficiaries would have been 28 percentage points higher in the absence of benefit receipt. Their results do not point to a greater employment potential for those with musculoskeletal and mental disorders. Quite the opposite, their estimates suggest that the foregone employment rate for those with musculoskeletal disorders is of 17.3 percentage points, almost half the magnitude of that of those with mental disorders, cancer, cardiovascular, neurological and digestive disorders[3].

French and Song (2014) exploit the variation arising from being assigned to different DI case judges in the US and arrive at the opposite conclusion. Those with musculoskeletal disorders have a higher employment potential. The discrepancy with Maestas et al. (2013) could lie on the fact that applicants suffering from musculoskeletal disorders are found to be significantly more likely to fill in an appeal after a rejection. During the appealing process, applicants may decide to keep their earnings low to increase their chances of winning the appeal.

These studies compare the inflow of DI recipients who never effectively entered DI. The outflow of DI has received far less attention. To my knowledge, Moore (2014) provides the only evidence so far on labor responses to terminated

[3]The point estimates for these range between 32.2% for digestive disorders to 36.6% for mental illnesses.

DI benefits by diagnosis. He exploits the exogeneity arising from a policy change in the US scheme in 1997, under which drugs and alcohol addictions no longer qualified for DI. The employment response to DI withdrawal for individuals with medically verifiable conditions is 25 percentage points lower than for those with musculoskeletal, mental or alcohol problems.

Despite the extensive literature on the topic, this thesis provides new evidence on a very particular reform, which specifically tackles the return to employment of DI claimants. Assessing its effectiveness can yield informative lessons on the ability of the administration to evaluate the residual capacity of DI beneficiaries.

The rest of the thesis is organized as follows. Section 2 provides the institutional context and detailed information about the reform. Section 3 describes the data and the empirical strategy, and justifies its application. Section 4 presents the results of the reform, first presenting the aggregate results, and then analyzing the heterogeneity in response by age and diagnosis. Section 5 concludes.

2. Institutional Context

This section aims at putting the Disability Insurance into its institutional context at the time of the reform, focusing only on the elements which are relevant for the analysis. In a second part, I will explain in detail the reform under inspection.

2.1. The Dutch Disability Insurance

Before 2004, three schemes were in place: the WAO, targeted to workers between 30-65 years old, the Wajong, for young disabled, and the WAZ scheme for self-employed workers. Workers in the Netherlands could apply for disability benefit after one year of sick pay.[4] A particularity of the Dutch DI system was that both work-related injuries or occupational diseases and non work-related sicknesses could qualify a worker for the scheme.

The social benefit administration (UWV) assessed the degree of disability of the applicant (DD) and corresponding disability benefit level. To determine these characteristics, the 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. The DD was calculated as follows:

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

Where $\overline{\omega}$ is the average earnings of the 10 best paying jobs the applicant could still perform, and ω^p are her earnings prior to the onset of the sickness. If the 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. The wage replacement rate was also dependent on the DD calculated by the UWV. Table 3 presents an overview of the wage replacement rate per DD. To assess its generosity, note that low levels of impairment could qualify an individual for DI, with the particularity that the benefit sometimes acted as a wage complement, i.e the DI claimant could work and receive DI benefits simultaneously.

Level of Disability	Benefit as a % of the Daily Wage
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%

Table 1: Levels of Benefit

[4] During the year of sick pay, a system of employment protection prevails: employers must pay 70% of the previous earnings to the incapacitated worker as well as maintain her job position.

Once the benefit was granted, the recipient underwent a first reassessment after one year, then periodically every 5 years. However, a report from the Work and Income Inspectorate (IWI, 2004) indicates that often these reassessments were nothing more than a questionnaire that had to be returned; Effectively, it seems that these reassessments were only performed by demand of the claimant.[5]

2.2. The 2004 Reform

As of January 2004, the criteria employed to assess the degree of disability of new applicants to the scheme was tightened, and accepted DI claimants became part of a new DI scheme (the WIA). Additionally, 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 the new, 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

All those whose onset of sickness fell after January 2004 were absorbed by the WIA, a new disability law taking effect in 2006[6]. Among several other aspects (see Koning and Lindeboom, 2015), one of the differences with the WAO scheme was the eligibility criteria. In broad lines, there was more flexibility when evaluating the suitable jobs for the applicants, which increased the range of jobs considered when calculating the DD. Particularly, these changes consisted in considering full time jobs for part-time applicants, and matching claimants with jobs specifically requiring the use of Dutch and computer skills even if the applicant didn't meet the requisites[7]. Unless the medical condition of the applicant impeded it, jobs involving night shifts were also taken into consideration. Additionally, the DD was calculated by setting the maximal weekly hours worked to 38. Finally, only 3 instead of 10 highest paying suitable jobs were used to calculate the DD. As a result, for most cases the DD was lowered. It was anticipated that these changes would increase the rejection rate of new applicants by 4.6% (Van Deursen, Mulders (2009)).

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 current stock of 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. It was determined that only those recipients born after the 1st of July 1954 would be reassessed.

Initially, the plan was to reassess recipients by groups of age, from young to old, and to have finished by 2007. Effectively, there was quite some lack of structure in the reassessment process, and all eligible recipients had a positive probability to be reassessed thorough the reform, which lasted until mid-2009. However, the frequency of the reassessments by age indicates that the probability of younger recipients to be reassessed was higher than for older recipients at the early stages of 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[8]. Table 2 shows the outcomes of the reassessments. Overall, 32% of the reassessed recipients exited DI, i.e their DD fell below 15%. Among initially fully disabled (DD 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 40% - mainly the fully disabled- had their DD unchanged.

At the end of 2005, the share of younger individuals reassessed was larger, so one would expect the outcomes from table 2 to become less extreme over time, under the assumption that older individuals may be more disabled.

[5] The IWI report (2004) shows that the reassessments only induced 18 percent of the total outflow from DI.

[6] The new disability law took effect two years after the eligibility criteria became effective, because as part of the new scheme the period of sick pay was extended from one to two years.

[7] The underlying motive is that the claimant should be able to acquire these skills within six months.

[8] Unfortunately, such detailed level of information was only disclosed 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.

Table 2: 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)

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.[9] To ensure the validity of the estimation strategy, this cohort will not be used in the analysis below.

Table 3 provides an overview of the final outcomes of the reassessment process. A sizable 20% of the individuals did leave DI because of the reassessment reform, and 12% experienced a decrease in their disability benefit while remaining in DI.

Table 3: 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: Uitvoeringsinstituut Werknemersverzekeringen (2009)

3. Data and Empirical strategy

3.1. Data Sources and Sample Selection

I use administrative data maintained by Statistics Netherlands. The full 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, labor market and disability insurance information is gathered from the tax authority and social insurance records. I make use of the information on employment spells from the tax authority records to construct the outcome variables of interest. The social security records contain detailed information on the disability spells, main and secondary 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, I will focus only on WAO recipients[10] Each of the different sources contains an individual identification number (RIN-code) which allows me to merge datasets at the individual level. All data is organized in yearly records, and is available

[9] Particularly, 25,000 cases were revised (Staatsblad 2007, 324). That is: out of 53,000 recipients between 45 and 50 years old, 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.

[10] WIA recipients are omitted because they were not part of the reassessment wave. Self-employment data does not yield such a clear measure, so WAZ recipients are also omitted. Finally, Wajong recipients have never worked before, so re-employment effects become difficult. The WAO scheme, in any case, encompasses 80% of DI recipients in the Netherlands.

from 2000 until 2009.[11]

I limit and modify the data employed for the analysis in several directions. First, I re-organize the datasets on a monthly basis. I am able to do so, since I have precise information on the employment and disability spells of each individual; I can infer whether the individual was working or in DI in every month, for every year. Second, I select these individuals who were in DI at the beginning of their observation period (2000 or 2003), and follow them thorough, so that I effectively work with a balanced panel. The only source of attrition in the sample used is due to death. Following the previous discussion, I drop all individuals aged in between 45 to 49 in 2001 and in 2004.[12] This leaves me a sample with individuals aged in between 30 to 45 and 50 or above in 2001 and 2004. With this definition, there will be an overlap in the classification of individuals to the 2001 group or the 2004 group: all individuals aged in between 33 to 44 (and above 53) years old in 2004 were aged in between 30 to 44 (above 50) in 2001. To solve this issue, I treat same individuals as different observations on the basis of their age in 2001 and 2004, and the time period thorough which they are followed. I categorize them through the indicator variable *ID*.

Table 4: Descriptive Statistics

		Age in July 2001	Age in July 2004
Age Group	Variable	Mean	Mean
30 to 36	N	64753	65267
	Main Diagnostic (%)		
	- General	10.8	8.87
	- Cardiovascular	0.97	1.06
	- Musculoskeletal	17.05	20.36
	- Psychological	44.36	49.83
37 to 45	N	141929	143113
	Main Diagnostic (%)		
	- General	17.05	11.46
	- Cardiovascular	1.68	1.93
	- Musculoskeletal	22.76	24.64
	- Psychological	37.84	43.82
50 to 53	N	78753	79245
	Main Diagnostic (%)		
	- General	18.47	14.9
	- Cardiovascular	4.49	4.44
	- Musculoskeletal	29.94	30.16
	- Psychological	29.9	34.5

Although there are 14 different categories of diagnosis, I report only the proportions for endocrine, cardiovascular, musculoskeletal and mental disorders. The two latter are of particular interest, being labeled as "subjective" health conditions. The two first categories are chosen because they are based on "hard" medical evidence, and they had amongst the highest proportions in the sample, ensuring the robustness of the analysis on these medical conditions. Table 4 shows the proportion of recipients with these diagnoses. Psychological diseases capture by far the largest proportion of DI recipients for younger groups, and he proportion is monotonically decreasing with age. Instead, the share of musculoskeletal disorders increases with age, reaching parity with the share of mental illnesses for the oldest groups. Those developments are common to most OECD countries (OECD, 2015).

[11] Although the data was initially available for 1999, because of an update in the social security records at the time of the writing, the files containing the yearly DI information of 1999 became obsolete. I have decided to leave this information out, given the differences in the definitions of the old and new versions of the data. Additionally, because of the empirical strategy used in this paper, data from 2007 onwards will not be used.

[12] Although they were eligible for reassessment only until February 2007. Although the observation period stops in 2006, so that this group of age could be considered as being treated by the reform, I follow the conservative approach not to use it in the analysis for two main reasons. (1) The frequency of reassessments among these individuals was significantly lower than for other age groups, making it difficult to identify treatment effects. (2) There was strong opposition against reassessing this groups of age since the beginning of the reform, which may have resulted in a different treatment than for other groups of age. For example, cases in which the claimant aged in between 45 to 50 would clearly no longer classify for DI could have been prioritized.

3.2. Empirical Approach

I use a differences-in-differences framework to identify the average impact of the reform on the recipients eligible for reassessment. One simple difference-in-differences would compare changes in outcomes for the treated group aged below 45 years on July 1 2004 who were eligible for reassessment with changes in a control group aged 50-53 at that time. But since one expects older DI beneficiaries to have a lower probability of return to work than younger recipients, the common trend assumption is not plausible. An alternative DID would compare the changes experienced by the same treatment group with those of a control group defined to be of the same age as the treatment group but in a period prior to the reform. But in the presence of time specific economy-wide changes, the common trend assumption required for this approach to identify the effect of the reform is also doomed to fail.

The strategy followed here combines the two aforementioned counterfactual groups in a trend adjusted triple differences in the spirit of Bell et al. (1999). In short, this approach uses four groups to identify the treatment effect. The treatment group and the control group are defined with respect to age in 2004 and are followed over the period 2003-2006. Another two groups are identified such that there ages in 2001 correspond to the ages of the treatment and control groups in 2004. These two additional groups, which will be used to identify age-specific trends, are followed over the period 2000-2003. The individuals in each group are selected on the basis of being in DI at the beginning of the period over which they are to be followed (2000 or 2003). Whereas comparison of the treatment group with the older control group during the same time period (2003-2006) is used to control for time-specific trends common to both of these groups, the two additional groups observed in the earlier period are used to correct for age-specific trends.

The selection of the observation time period prior to the reform is conditioned on proximity in terms of its economic and DI conditions to the period of the reform, as will be explained in detail in the next subsection.

The regression specification of the first differences-in-differences is:

$$Y_{igt} = \kappa T_t + \delta D_g T_t + \gamma_g m_t + \mu_i + \epsilon_{igt} \tag{1}$$

Where Y_{igt} is the employment outcome of interest for a claimant i, of age group g, at time t. μ_i is an individual unobservable term, $\gamma_g m_t$ is an age specific linear time trend, with $m_{t+1} = m_t + 1$, and ϵ_{igt} is an idiosyncratic error term (iid over i, g and t). D_g takes value 1 if the age of the individual qualifies him to be reassessed, and T_t is a pre/post reform dummy that takes value 1 at the onset of the reassessment process, i.e 1st of October 2004. The interaction between the two takes value 1 in the post-reform period for the treated claimants. The parameter δ captures the impact of the reassessment reform, hence is the parameter of interest.

In a two-period framework, I set $t_0 = 2003$ and $t_1 = 2006$, such that t_0 and t_1 are the pre- and post-reform time periods, respectively. Suppose I label the control group C, and the treatment group T. Then, the first difference for treated beneficiaries becomes:

$$\begin{aligned} \tilde{Y}^T_{t_1} - \tilde{Y}^T_{t_0} &= \kappa + \delta + \gamma_Y (m_{t_1} - m_{t_0}) \\ \text{where } \tilde{Y}^T_t &= E[Y_{igt} | D_g = 1, t = \{t_0; t_1\}] \end{aligned} \tag{2}$$

For the control group it is:

$$\begin{aligned} \tilde{Y}^C_{t_1} - \tilde{Y}^C_{t_0} &= \kappa + \gamma_O (m_{t_1} - m_{t_0}) \\ \text{where } \tilde{Y}^C_t &= E[Y_{igt} | D_g = 0, t = \{t_0; t_1\}] \end{aligned} \tag{3}$$

Taking differences-in-differences will identify the treatment effect δ under the assumption that $\gamma_Y = \gamma_O$. That is, the age specific component in the time trend is required to be from the same distribution in both groups. This is limiting, as previously explained, if we expect age to have a differential impact on the dynamics of employment and DI exit rates.

I follow Bell et al. (1999), and make use of two groups of individuals from periods prior to the reform to identify the age specific linear trends. In a two period set up, let $t^*_0 = 2000$ and $t^*_1 = 2003$, such that the time period of

observation prior to the reform is spanned. The model to be estimated becomes:

$$Y_{igt} = \kappa T_t + \delta D_g P_i T_t + \gamma_g m_t + \mu_i + \epsilon_{igt} \tag{4}$$

The indicator variable P_i takes value 1 for the groups followed from 2003 to 2006, and 0 for those followed from 2000 to 2003. Note that since D_g is defined relative to the age, those beneficiaries aged in between 30 to 45 both in 2001 and 2004 have $D_g = 1$.

Whereas the parameter κ captures the common dynamics in the post-reform period that is the same for both age groups, $\gamma_g m_t$ captures age specific differences in time trends. μ_i are again individual time invariant characteristics. Then, the treatment effect is captured by δ.

Note however, that the set of identifying assumptions changes from the DID situation. To see this, I label the individuals tracked from 2000 to 2003 Y for the young, and O for the older. Let the first differences in a two period set up of the four groups in (4) be

$$\tilde{Y}^T_{t_1} - \tilde{Y}^T_{t_0} = \kappa + \delta + \gamma_Y(m_{t_1} - m_{t_0}) \tag{5}$$
$$\tilde{Y}^C_{t_1} - \tilde{Y}^C_{t_0} = \kappa + \gamma_O(m_{t_1} - m_{t_0}) \tag{6}$$
$$\tilde{Y}^Y_{t^*_1} - \tilde{Y}^Y_{t^*0} = \gamma_Y(m_{t^*_1} - m_{t^*_0}) \tag{7}$$
$$\tilde{Y}^O_{t^*_1} - \tilde{Y}^O_{t^*_0} = \gamma_O(m_{t^*_1} - m_{t^*_0}) \tag{8}$$

Where $\tilde{Y}$ is again the expectation of the outcome of interest. The triple differences is then:

$$[(\tilde{Y}^T_{it_1} - \tilde{Y}^T_{it_0}) - (\tilde{Y}^C_{it_1} - \tilde{Y}^C_{it_0})] - [(\tilde{Y}^Y_{it^*_1} - \tilde{Y}^Y_{it^*_0}) - (\tilde{Y}^O_{it^*_1} - \tilde{Y}^O_{it^*_0})] = \delta + (\gamma_Y - \gamma_O)(m_{t_1} - m_{t_0}) - (\gamma_Y - \gamma_O)(m_{t^*_1} - m_{t^*_0}) \tag{9}$$

The identification of the treatment effect will only hold if $(\gamma_Y - \gamma_O)(m_{t_1} - m_{t_0}) = (\gamma_Y - \gamma_O)(m_{t^*_1} - m_{t^*_0})$. With age-specific linear time trends, this equality imposes that the time span of observation is the same for all groups. Particularly, it imposes (1) same length of observation periods for all groups, and (2) a linear time trend for m_t that allows the comparison between groups from different periods.

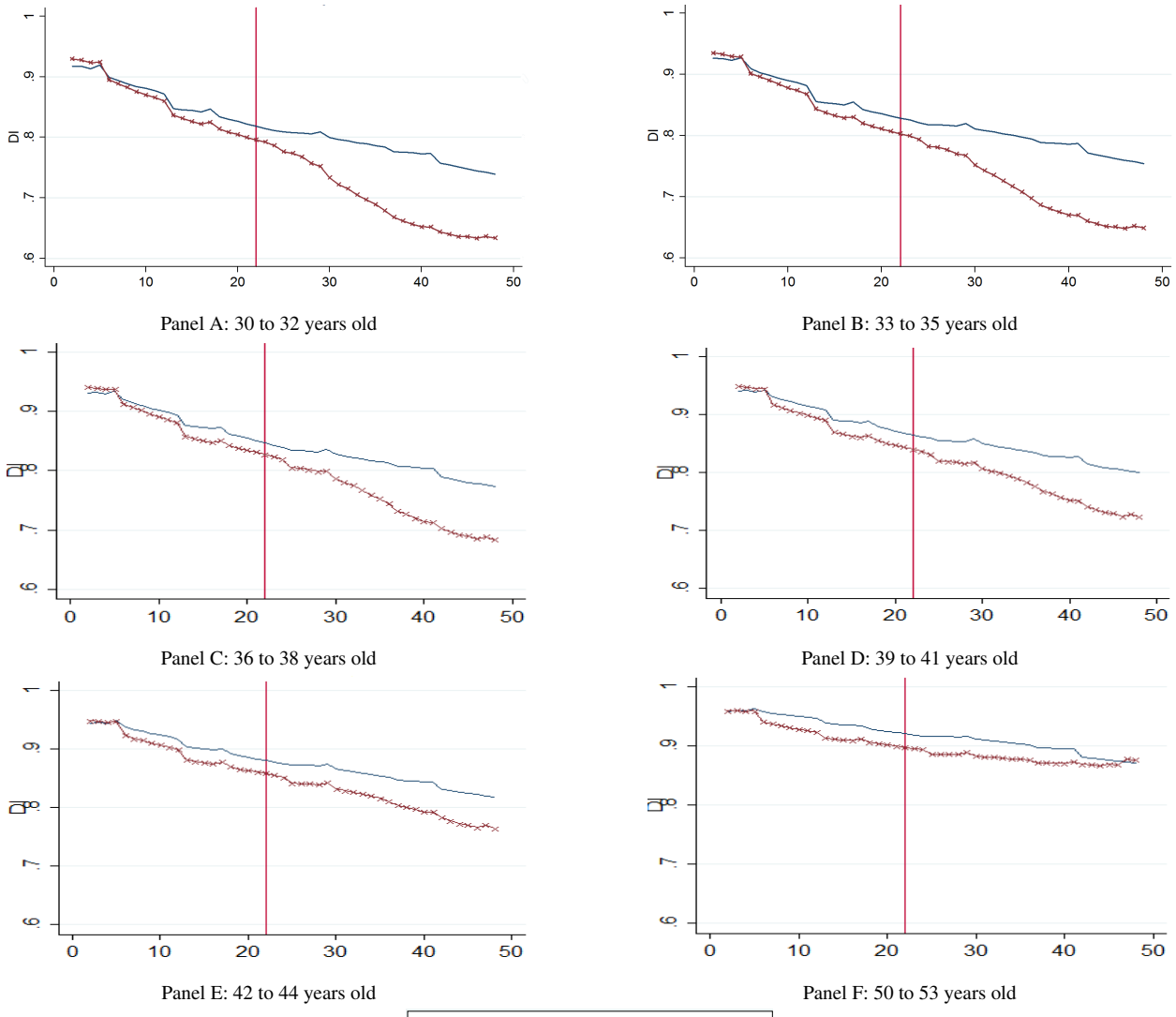
The identifying assumption is that the time trends can be modelled by a linear age-specific term, allowing for common dynamics between individuals observed in a same time period to affect the outcome of interest. This imposes a limitation if there are "cohort effects", for instance[13]. In the next subsection, I argue about the validity of these assumptions, as well as present a graphical analysis of the impact of the reform

3.3. Graphical Analysis

A crucial factor when selecting the observation periods is that the economic environment should not affect differently the trends of older and younger claimants in one period than in the other. For instance, if the economic environment was dramatically different during the period prior to the reform than during the reform period, chances are that the trend differentials between ages are not constant across periods. It is then a good approach to select a period as close as possible to the reform period. Selecting DI claimants in January 2000, and tracking them until December 2003, ensures a sufficient number of observations to evaluate the plausibility of the assumptions of the model. I will to so by performing a graphical inspection of the trends in the outcomes of interest for the four groups employed, in their respective time periods of observation.

Figure 1 enables us to make a comparison on the evolution of the probability of being in DI across time periods, for different age groups of recipients. These horizontal axis in the graphs is a constructed time trend following the previous discussion, such that it takes value 0 in January 2003 and ranges up to 48 (December 2006) for the groups contemporaneous to the reform, and takes value 0 to 48 (January 2000 until December 2003) for the recipients selected from a time period prior to the reform. As the legend indicates, the crossed line depicts the evolution of the probability to remain in DI for the groups contemporaneous to the reform, selected in 2003. In panels A to E, the crossed line corresponds to the treated individuals; In panel F, it corresponds to the control group. The solid line depicts the trend

[13] If being 30 years old in 2001 would result in a different time trend than being 30 in 2004. This could be so if there are non-linearities in the relationship between age and time.



Panel A: 30 to 32 years old

Panel B: 33 to 35 years old

Panel C: 36 to 38 years old

Panel D: 39 to 41 years old

Panel E: 42 to 44 years old

Panel F: 50 to 53 years old

- The age of the claimants is measured in July 2001 for those observed in 2000-2003, and July 2004 for those observed in 2003-2006
- All claimants are in DI at the beginning of their respective observation period

Figure 1: Probability to remain in DI

for the recipients selected in 2000. To maintain the symmetry with the groups contemporaneous to the reform, their age is measured in reference to July 2001. The vertical line marks the month that corresponds to the beginning of the reassessment process, i.e October 2004, for observations followed over the 2003-06 period.

At first sight, note that the evolution in the probability to remain in DI seems to be well approximated by linear trends. This gives us confidence in the linear specification adopted above. The trends in the probability to remain in DI starting in 2000 and in 2003 seem to match quite closely in the pre-reform periods. However, there is a noticeable gap appearing between the two trends after the first year of observation in panels A-E. This could be an issue if we were using those recipients in 2000 as a control group for their treated peers in 2003. Indeed, assuming common trends between these groups would imply underestimating the probability to leave DI for the treated, resulting in a grossly overestimated impact of the reform. On the other hand, we also find that older individuals experience flatter trends in the probability to remain in DI. This implies that using the older group that is not reassessed as a control group without correcting for trend differences due to age would also overestimate the impact of the reform.

Note that the beginning of the reassessment process did not have an immediate effect on the probability to remain in DI. This can be explained by the strong opposition against the reassessment process that arose at the onset of the reform. The magnitude was such that during the first months of the reform, very little reassessments took place (UWV, 2005). The magnitude of the sudden fall in the probability to remain in DI in the post period strongly depends on the age of the recipients analyzed. Recipients aged in between 30 to 38 experience a sharper decline than these aged in between 39 to 44. Recall that the re-evaluated degree of disability depended to a larger extent on the job market opportunities of the reassessed individual. The less pronounced drop in the probability to remain in DI for older reassessed recipients could be a consequence of their greater unfitness to the labor market. We do not observe a change in the trend after the implementation of the reform non-reassessed recipients (Panel F). This observation reinforces the idea that the reform did not have spillover effects on non-eligible individuals.

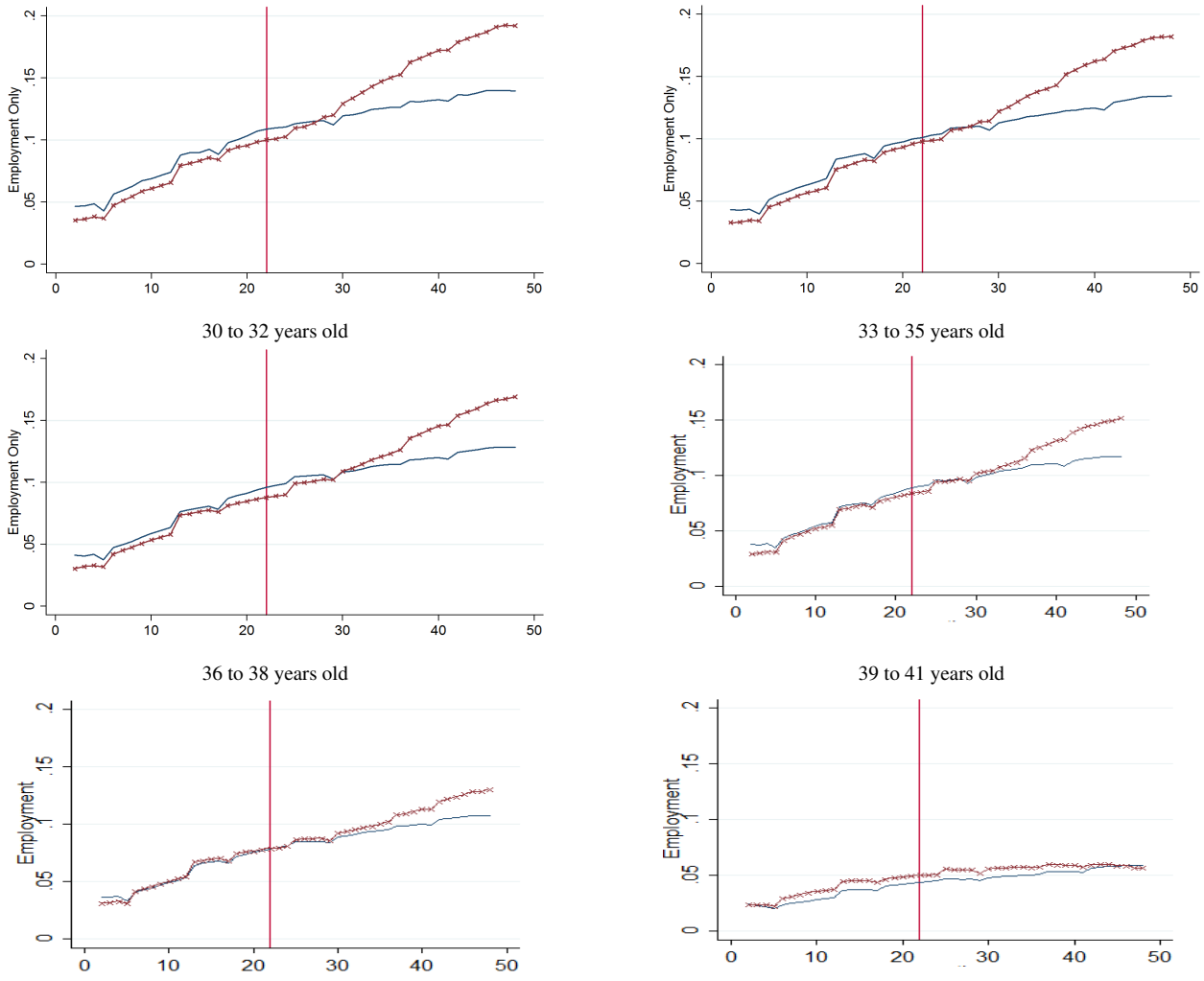
The set of panels in Figure 2 show the trends in the probability to be in employment only. This variable will be the employment outcome of interest thorough the paper. Note that total employment is an aggregated figure of the probability of being in DI and employment, and the probability of being employed and not in DI. Focusing on the latter gives a better picture of the impact of the reform on the individuals that were rejected from DI. The trends match closely across periods for the pre-reform period, particularly for the age groups in between 39 to 44. Note that by construction, the probability of being employed only is increasing over time, as it was zero at the beginning of the period. Still, it seems clear that the reform had employment effects. There is a sharp increase in the probability of being employed only for the treated individuals, starting a few months after the reassessment process started. This delayed employment response coincides with the delayed DI outflow boost. The fact that the delay is visible in both series gives confidence in that there were no sizable anticipation effects: it does not appear that recipients, expecting to be rejected from DI, left DI before the reform started and instead increased their labor market participation.

In the appendix, I show the disaggregated probability to remain in DI, separating by DI only and DI and employment.

The evidence presented until now has shown that the reform had a sizable impact on the probability to remain in DI and of employment. Although the identifying assumption of no cohort effects is not testable, the closeness of the pre-reforms trends gives us confidence on its plausibility. In what follows, I will present the results from the estimation regressions previously derived.

4. Results

This section is organized in three different parts. First, I evaluate the effectiveness of the reform in the aggregate. Due to computational limitations, these estimates make use of yearly data. This may result in a bias in the estimates, due to a potential overestimation of the employment in the sample. To overcome computational limitations as well as to explore the heterogeneity in the treatment impact, I perform the analysis by age, closely following the graphical analysis previously presented. Finally, I present the results by allowing for heterogeneity in response by main diagnosis.



- The age of the claimants is measured in July 2001 for those observed in 2000-2003, and July 2004 for those observed in 2003-2006
- All claimants are in DI at the beginning of their respective observation period

Figure 2: Probability to be Employed Only

4.1. Average impact of the reform

The baseline results are obtained from the estimation of (4). Its empirical counterpart is defined as

$$Y_{igt} = \kappa T_t + \delta D_g P_i T_t + \gamma_1 \sum_{t=1}^{3} time_t + \gamma_2 D_g \sum_{t=1}^{3} time_t + \mu_i + \epsilon_{igt} \tag{10}$$

A linear time trend is included to capture $\gamma_g m_t$: $time_t$ is a linear yearly indicator, taking value 0 in 2000 for the groups selected in 2000, and ranging until 3 in 2003. Similarly, $time_t$ takes value 0 for 2003 for the groups contemporaneous to the reform, and ranges until 3 in 2006. Including an interaction term between D_g and the time trend, we allow time trends to differ depending on whether the recipient belongs to a group of young age (i.e aged in between 30 to 45) or of old age (i.e. older than 50 years of age). Because of the yearly nature of the data, the treatment needs to defined on a specific year. One approach is to set the treatment in 2004, with the underlying assumption that the full impact of the reform has been reached at that time (Mora & Reggio, 2012). Given that the initial stages of the reform did not entail a large number of reassessments, this approach may lead to a serious underestimation of the treatment effect. A different approach is to set the reform to take place in 2005. If there are anticipation effects, doing so may provide a lower bound of the actual impact of the reform. I choose the latter, and interpret my results as a low bound of the impact of the reform.[14].I then need to redefine T_t as being 0 for the pre-reform years, 2003-2004, and 1 otherwise. Specification (10) is then estimated through individual fixed effects, with standard errors clustered at the *ID* level (Donald and Lang (2001))[15].

Table 5 shows the baseline estimates, resulting from the estimation of (10). Column 1 includes all observations in the sample. The results show a negative significant impact of the reform on the probability to remain in DI: recipients eligible for reassessment experience on average a decrease in 3.5 percentage points in their probability to remain in DI.

Table 5: Baseline Estimates

	Full sample	DI	DI and Work
DI	-3.5*** (0.14)	-3.87*** (0.15)	-2.95*** (0.27)
Employment	3.66*** (0.12)	2.62*** (0.09)	5.43*** (0.26)
Observations	572229	382411	189818

Significance levels: *** 1%, ** 5%, * 10%. Standard errors reported in parenthesis.
Estimates are from a linear probability model, x100.
Estimation through *ID* fixed effects.
Clustered s.e. at the *ID* level.

The employment outcome reported captures the probability of being employed and not in DI, so that there is no overlap in the definition of Employment and DI in Table 5. Because initially all observations are in DI, the increase in the probability of being employed can only come at expenses of a decrease in the probability of remaining in DI. However, the estimates show larger percentage increases in employment than percentage drops in DI. This surprising result may be due to the yearly nature of the data. Any individual having had a job while not being in DI for at least one day during a year is accounted as employed during that year; this may result in an overestimation of the amount of employment while out of DI. Since the category of DI includes as well DI claimants who work, there is a risk of overestimating the amount of observations in this category[16]. These two biases may result in an underestimation of the impact of the reform on DI, and an overestimation of the impact of the reform on employment, potentially explaining why the percentage point increase in employment is larger than the percentage point decrease in DI.

[14]The DI outflow estimates when setting the treatment in 2004 are around 0.5 percentage points smaller in absolute value than when setting the treatment in 2005. This confirms that the small number of reassessments taking place in 2004 generates a more severe underestimation of the impact of the reform than omitting anticipation effects.

[15]Estimating the models using clustered standard errors per individual yields entirely robust estimates to those at the *ID* level.

[16]Because the movements in and out of DI for a same individual within a year are much less recurrent than for employment, it is less likely that there is an overestimation of DI claimants within a year.

Columns 2 and 3 decompose the sample depending on the initial situation of the beneficiary. Column 2 shows those who were in DI only at the beginning of their observation period, and column 3 those initially in DI and employed at the same time. The probability to remain in DI drops more for these individuals who were initially only in DI: they experience a decrease in 3.9 percentage points in the probability to remain in DI, whereas these initially in DI and employed experience a 3 percentage point decrease. The employment estimates for those initially working and in DI point to a 5.4 percentage point increase in employment, 1.4 percentage points larger than the increase in DI outflow. This improbable finding seems to corroborate the overestimation of these employed and in DI. This is because those initially in DI and working have a higher probability of being in DI and working at the time of the reform than those initially in DI only. If this is true and there is an overestimation of those in DI and working, their DI estimate would suffer from a stronger upward bias than those initially in DI only.

Overall, although the estimates show the success for the reform, it is not possible to omit the fact that there may be suffering some serious bias. The next subsection sheds light on this issue.

4.2. Age Heterogeneity

To allow for comparison with the baseline estimates, I first present the heterogeneous impact of the reform per age groups on a yearly basis, maintaining the same treatment definition (i.e treatment defined in 2005). Effectively, I estimate equation (10) per age, by grouping treated individuals into five different age groups: those aged between 30 and 32, 33 and 35, 36 and 38, 39 and 41 and 42 and 44. Individuals from the time period prior to the reform are selected so to match those age groups. The two groups of older individuals aged between 50 to 53, are used in every estimation[17]. Fixed effects and clustering of the standard errors takes place at the individual level. Table 6 shows the results from the estimation for different groups of age, and reports the aggregate effects.

Table 6: Effect of the reform by age. Yearly data.

	Estimates		
Age Group	DI	Employment	Individuals
30 to 32	-5.81***	5.59***	213486
	(0.28)	(0.25)	
33 to 35	-5.41***	4.99***	232406
	(0.24)	(0.21)	
36 to 38	-3.43***	4.03***	241618
	(0.23)	(0.19)	
39 to 41	-2.78***	3.09***	252936
	(0.21)	(0.18)	
42 to 44	-1.47***	1.84***	264035
	(0.19)	(0.16)	
Aggregate	-3.5***	3.66***	572229
	(0.14)	(0.12)	

Significance levels: *** 1%, ** 5%, * 10%.
Standard errors reported in parenthesis.
Estimates are from a linear probability model, x100
Estimation through *ID* fixed effects.
Clustered s.e. at the *ID* level.

As already expected from the graphical analysis, the impact of the reform decreases with age: the reform affected more heavily the probability to remain in DI for the youngest cohorts. This must not be directly interpreted as the youngest cohorts being particularly affected by the reform, but one must acknowledge the fact that at the time of observation younger individuals were being reassessed with higher frequency. Overall, it seems that the probability of being employed increases for all groups of age, ranging from 1.84 to 5.59 percent. The employment estimates for

[17]The grouping of individuals by three years only avoids any overlap in the group definition of each individual.

those aged in between 36 to 44 are larger in absolute value than the DI estimates. To shed light on the amount of overestimation of employment from using yearly data, I now move to explore the effects of the reform by age with monthly data.

I estimate again (10), where $time_t$ is now month-year linear time trend, as defined in section 3.3. That is, taking value 0 for January 2000 and 48 for December 2003 for the group selected in 2000, and ranging from 0 for January 2003 to 48 for December 2003 for the group selected in 2003. The time of the reform is now set to be October 2004. Table 7 shows the results from the specification (10), for the probability to remain in DI and the employment response. Column 1 shows the estimates for the sample of all individuals in the age group under analysis. First, note that there seems to be a significant difference in the point estimates from the yearly and monthly specifications. Whereas yearly estimates for the probability of remaining in DI are lower than its monthly counterparts, the estimates of the probability to be employed only are larger. This finding is entirely aligned with the hypothesis that the DI and employment estimates suffer from a positive bias when using yearly data.

Table 7: Effect of the reform by age. Monthly Data.

	Full Sample	DI	DI and Work
30 to 32			
DI	-7.06*** (0.25)	-7.33*** (0.28)	-6.98*** (0.49)
Employment	3.2*** (0.2)	2.1*** (0.18)	5.3*** (0.46)
Observations	213486	141601	71885
33 to 35			
DI	-6.6*** (0.2)	-7.1*** (0.24)	-5.9*** (0.43)
Employment	2.9*** (0.2)	1.95*** (0.21)	4.97*** (0.401)
Observations	232406	154540	77866
36 to 38			
DI	-4.8*** (0.2)	-5.3*** (0.2)	-3.78*** (0.4)
Employment	1.9*** (0.2)	1.6*** (0.2)	2.73*** (0.38)
Observations	241618	160466	81152
39 to 41			
DI	-3.8*** (0.2)	-4.03*** (0.2)	-3.09*** (0.37)
Employment	1.6*** (0.1)	1.03*** (0.11)	2.42*** (0.35)
Observations	252936	167666	85270
42 to 44			
DI	-2.2*** (0.2)	-1.99*** (0.18)	-2.87*** (0.34)
Employment	0.90*** (0.2)	0.516*** (0.09)	1.99*** (0.32)
Observations	264035	173470	90565

Significance levels: *** 1%, ** 5%, * 10%. Standard errors reported in parenthesis.
Estimates are from a linear probability model, x100.
Estimation through individual fixed effects. S.e. clustered at the individual level.

Column 1 in Table 7 shows again that younger groups of age experience larger DI outflow and employment increases than older claimants. This fact can be reconciled both with younger claimants being more frequently re-assessed at the time of observation, and younger claimants being fitter for the labor market according to the administration and in reality. The percentage point increase in employment is less than half the size that of the DI outflow.

Despite correcting from the bias from the use of yearly data, the impact of the reform on DI remains larger on those initially in DI than on those initially working and in DI. It seems that the administration evaluated those initially in DI to be fitter for the labor market, resulting higher rejection rates from DI among these. The employment response is significantly larger for those initially in DI and working. This can be rationalized if labor market frictions and discrimination are sizable among disabled individuals. One can argue that those who were initially employed and in DI experienced less frictions as they were more frequently employed at the time of the reform. Another hypothesis is that these who were initially in DI only had a lower value for work. If this is so, despite being rejected DI, they may still prefer not to work. Finally, it could be that the administration is making more type II errors on awarding DI benefits to those initially in DI only.

Table 8 shows the estimates of the probability of being in DI only, and in DI and working. Overall, both the probability to remain in DI only and in DI and employment fall. This implies that there doesn't seem to be an increase in the labor force participation among the individuals who remain in DI. Rather, the reform fostered the participation

of the labor force for individuals who effectively left DI.

Table 8: Decomposition of the estimates by groups of age

	Full Sample	DI	DI and Working
30 to 32			
DI Only	-6*** (0.3)	-6.95*** (0.31)	-2.68*** (0.4)
DI & Employment	-1.2*** (0.2)	-0.39** (0.19)	-4.3*** (0.54)
33 to 35			
DI Only	-5.3*** (0.2)	-6.45** (0.268)	-2.37*** (0.43)
DI & Employment	-1.27*** (0.2)	-0.61*** (0.17)	-3.57*** (0.49)
36 to 38			
DI Only	-4.2*** (0.2)	-5*** (0.2)	-1.4*** (0.36)
DI & Employment	-5.3** (0.2)	-0.3 (0.2)	-2.38*** (0.46)
39 to 41			
DI Only	-3.4*** (0.2)	-4.09*** (0.23)	-0.8** (0.34)
DI & Employment	-0.4** (0.2)	0.6 (0.15)	-2.24** (0.44)
42 to 44			
DI Only	-1.8*** (0.2)	-1.8*** (0.21)	-0.31 (0.32)
DI & Employment	-0.4** (0.2)	-0.15 (0.14)	-2.56*** (0.41)

Significance levels: *** 1%, ** 5%, * 10%. Standard errors reported in parenthesis.
Estimates are from a linear probability model, x100.
Estimation through individual fixed effects. S.e. clustered at the individual level.

Estimating flexibly the reform can shed light on whether younger individuals were found to be fitter for the labor market under the stricter criteria, or if instead their DI outflow rises more heavily because their frequency of reassessment was higher.

To flexibly estimate the effect of the reform, I specify the set R as containing the post-reform time periods. With monthly data, this means all months after October 2004. The specification can be then rewritten as

$$Y_{igt} = \kappa \sum_{t \in R} time_t + \delta D_g P_i \sum_{t \in R} time_t + \gamma_1 \sum_{t=2}^{48} time_t + \gamma_2 D_g \sum_{t=2}^{48} time_t + \mu_i + \epsilon_{igt} \tag{11}$$

Equation (11) is estimated for the aggregate sample, and per age group g. The graphs in figures 3 and 4 depict the estimated treatment coefficient for the probability of being in DI and in the labor force, for each of the five age groups analyzed. One should interpret these figures as follows. The origin corresponds to the estimate of the outcome variable of interest in October 2004. Each of the following points correspond to the treatment effect in the subsequent months, ranging until December 2006. The solid line plots the point estimates of δ from (11), and the dashed lines the confidence intervals at 95%.

As suspected, the impact of the reform becomes larger over time, across all age groups: Figure 3 shows a decreasing trend on the point estimates of the probability of being in DI. The evolution of estimates for the two youngest groups flattens at around month 21 (May 2006); This plateauing is not suggested in any of the other graphs. I interpret this as the reform going through the reassessment process among younger individuals faster than among older individuals, so that the estimates of the impact of the reform reach an equilibrium value. It seems that ultimately, the impact of the reform converges to a same value for those aged between 30 to 32 and those between 33 to 35 years old, at around 12 percentage points. It is not possible to compare the impact of the reform across the other age groups, given that the full impact of the reform has not been reached.

The point estimates for the probability of being employed experience a sluggish start at the onset of the reassessment process, but quickly increase over time. For the two youngest groups of age, the increase in the probability reaches above 6 percentage points by the end of the observation period, compared to the 3.2 and 2.9 percentage points obtained from the discrete treatment specification. It is evident from the flexible specification that the gap between the increase in DI outflow and employment decreases as the reform progresses. This result does not contradict the hypothesis that the labor market frictions prevented employment rates to rise immediately as the reassessments started. On the other hand, the gap is still sizable, even among young individuals where the full effect of the reassessment

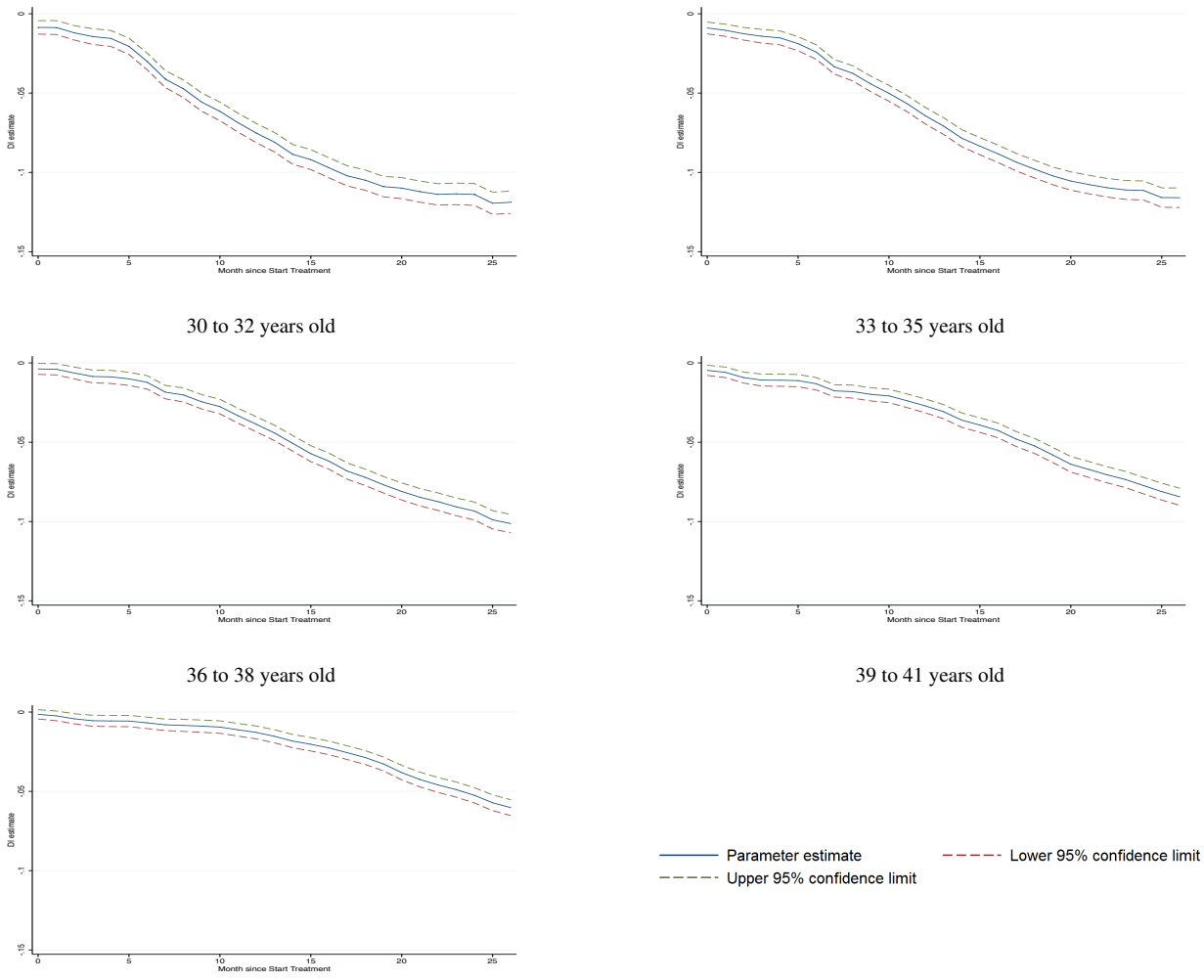
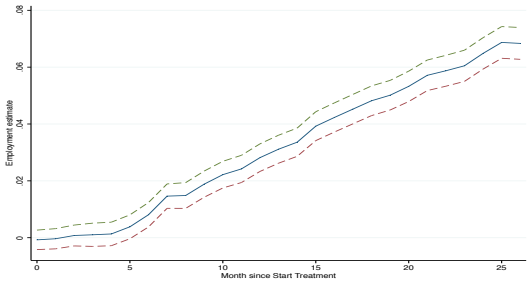


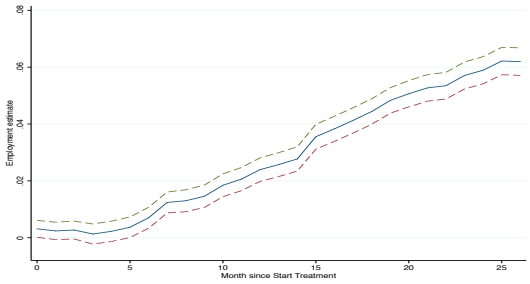
Figure 3: Flexible Specification Estimates: Probability to be in DI

process seems to have been reached.

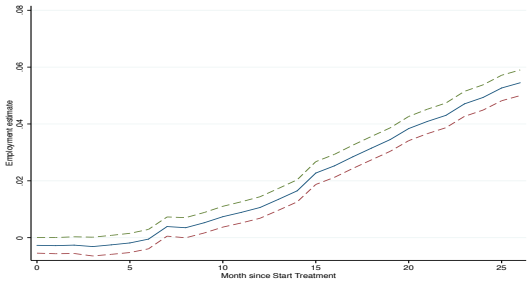
So far, we have seen that the rise in DI outflow generated by the reform is not accompanied by an increase in the employment rate of the same degree. An analysis of the dynamics of the employment response, however, shows that the gap between both responses is decreasing as the reform proceeds. I have interpreted this as highlighting the evidence that labor market frictions have prevented to reach the full employment potential of the sample of treated individuals immediately at the onset of the reassessment. Still, the remaining gap may be an indication of either type II errors from the administration when granting DI benefits, or the unwillingness to work of the treated claimants.



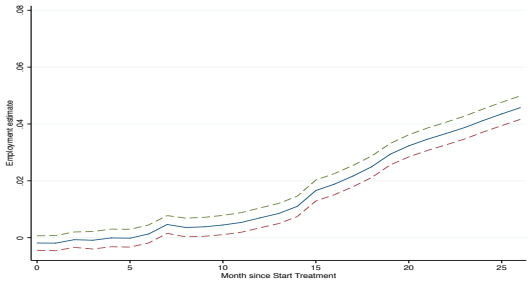
30 to 32 years old



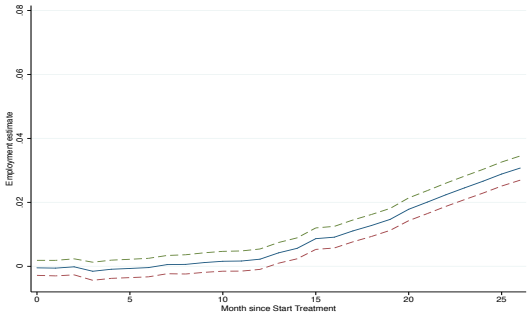
33 to 35 years old



36 to 38 years old



39 to 41 years old



42 to 44 years old

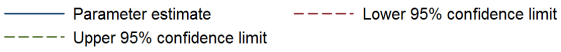


Figure 4: Flexible Specification Estimates: Probability to be Employed

4.3. Main Diagnosis Heterogeneity

In this subsection, I explore the heterogeneity across individuals with different health conditions. Technically, the new eligibility rules employed in the reform did not discriminate between health conditions. If this is true, differentials in the exit rates from DI across health conditions should reflect the differentials in the work impediment they impose. The residual employment capacity is often found to be larger for less verifiable health conditions, such as musculoskleletal or psychological illnesses. However, this is often confounded with the fact that able-to-work DI claimants with low work value may use less verifiable health conditions to shirk work. When it comes to actual work impediment imposed by "subjective" health conditions, it is difficult to argue that claimants suffering from back pain or depression are more able to work than those suffering from diabetes. It becomes then difficult to predict the relative responses of employment by medical condition.

I estimate equation (10), separately by health condition. The first row in table 9 shows the results of the estimations for four different diagnostics: endocrine diseases, cardiovascular, musculoskeletal disorders and psychological diseases.[18] Musculoskeletal disorders have experienced the highest rose in their DI outflow, 7 percentage points, followed by those suffering from mental illnesses, those with endocrine diseases, and those with heart diseases. According to the previous reasoning, the exit rates from DI are reflecting the evaluated residual employment capacity: the administration has considered claimants suffering from musculoskeletal disorders to be more able to work than those suffering from mental disorders or endocrine diseases. The employment responses are the largest for musculoskeletal disorders, with an increase in 3.66 percentage points, followed by mental (1.78 pp) , endocrine (1.21 pp) and cardiovascular diseases (1.16 pp). The ratios between the responses in employment and DI generated by the reform can give us an indication on accurate was the evaluation of the administration for different health conditions. Interestingly, those suffering with heart diseases seem to have experienced the most accurate evaluation of their prospective employment, as their ratio is the largest. The employment response for those with musculoskeletal disorders is of half the size of the rise in DI outflow, and it is only 40 percent for those with mental and endocrine disorders.

Rows 2 to 5 show the estimation of (10), both by health condition and age. The first column presents the results of the sample aggregated by medical condition, already reported in Table 7. Claimants suffering from musculoskeletal disorders experience higher rises in the DI exit rate and employment than their age group average. The degree to which their employment response accompanies rises in DI exit rates is strongly decreasing with age. This development does not hold for those with mental disorders, nor it does for those with heart and endocrine diseases, which experience constant ratios between DI exit and employment rises. The age-dependency of the gap for those suffering from musculoskeletal disorders could indicate that older individuals suffer more from labor market frictions and discrimination, from type II errors in their DI award, or have less willingness to work than their younger peers.

In the appendix, table 11 presents a decomposition of the measures of the probability of being in DI and employed, for the four different main diagnoses analyzed.

[18] Note that the full sample is not included for comparison, given the differences arising between yearly and monthly data. Because the sample is here divided per disability cause, it is possible to obtain estimates using monthly data.

Table 9: Treatment Effects per Group of Age and Main Diagnosis

	Full Sample	Endocrine	Cardio.	Musc.	Mental
Aggregate					
DI		-3*** (0.3)	-2.12*** (0.78)	-7.06*** (0.25)	-4.4*** (0.2)
Employment		1.21*** (0.22)	1.16*** (0.56)	3.66*** (0.19)	1.78*** (0.15)
Observations		95353	15523	154720	258592
30 to 32					
DI	-7.06*** (0.25)	-5.32*** (0.7)	-7.33*** (2.4)	-14.7*** (0.63)	-6.6*** (0.36)
Employment	3.2*** (0.2)	0.1 (0.67)	-3.26* (1.75)	8.34*** (0.55)	2.37*** (0.28)
Observations	213486	38433	8704	63865	86986
33 to 35					
DI	-6.6*** (0.2)	-6.26*** (0.68)	-4.66** (1.9)	-12.6*** (0.5)	-6.14*** (0.3)
Employment	2.9*** (0.2)	-2 (0.58)	3.2** (1.52)	6.56*** (0.44)	2.34*** (0.25)
Observations	232406	40384	9051	77866	95883
36 to 38					
DI	-4.8*** (0.2)	-4.76*** (0.5)	-1.92 (1.5)	-7.7*** (0.4)	-4.8*** (0.3)
Employment	1.9*** (0.2)	0.4 (0.5)	1.45 (1.11)	3.7*** (0.35)	1.81*** (0.22)
Observations	241618	43386	9378	72577	98360
39 to 41					
DI	-3.8*** (0.2)	-2.21*** (0.46)	-3.16** (1.25)	-6.1*** (0.36)	-3.72*** (0.3)
Employment	1.6*** (0.1)	0.42 (0.43)	1.42 (1)	2.9*** (0.32)	1.36*** (0.21)
Observations	252936	46746	9944	77238	101454
42 to 44					
DI	-2.2*** (0.2)	-1.66*** (0.4)	-0.8 (1)	-3.8*** (0.32)	-2.4*** (0.26)
Employment	0.9*** (0.2)	0.12 (0.36)	-0.08 (0.81)	2.03*** (0.27)	1.13*** (0.21)
Observations	264035	51359	10672	81690	103190

Significance levels: *** 1%, ** 5%, * 10%. Standard errors reported in parenthesis.
Estimates are from a linear probability model, x100.
Column 1 Row 1 is left empty, since the aggregate effect for all medical conditions comes from a yearly specification.
Estimation through individual fixed effects. S.e. clustered at the individual level.

4.4. Robustness Tests

To assess the plausibility that the strategy followed identifies the impact of the reform, I perform robustness tests on the estimation of (10). To do so, I set a placebo reform in each month of the year 2003, from February to November, and analyze the response of the outcomes of interest. If the estimates from this placebo analysis are statistically significant, one could doubt about the validity of the empirical strategy in correcting for trend differences across groups prior to the reform. Referring to the identifying assumption previously derived, it could be that there are strong cohort effects which invalidate the results. Table 10 presents the outcomes for the all groups of age, where the placebo reform is set in month 2 (February 2003), month 5 (May 2003) and month 11 (November 2003)[19]. The placebo reforms do not have any impact in the probability to being in DI and employed across age groups. These results give confidence that the strategy followed captures the impact of the reform, rather than pre-existing trend differences between the groups employed.

[19]The results are consistent when setting the placebo reform in any other month between February and November.

Table 10: Placebo Reforms

	February	May	November
30 to 33			
DI	-0.05 (0.21)	0.012 (0.21)	-0.05 (0.18)
Employment	-0.173 (0.153)	-0.123 (0.148)	-0.86 (0.14)
33 to 36			
DI	0.09 (0.18)	0.17 (0.17)	0.04 (0.15)
Employment	-0.3* (0.15)	-0.1 (0.14)	-0.06 (0.15)
36 to 39			
DI	0.11 (0.17)	0.12 (0.16)	0.07 (0.14)
Employment	0.17 (0.14)	0.22 (0.13)	0.28 (0.136)
39 to 42			
DI	-0.12 (0.15)	-0.22 (0.15)	-0.08 (0.13)
Employment	-0.2 (0.13)	-0.2 (0.13)	-0.17 (0.12)
42 to 45			
DI	0.08 (0.14)	0.09 (0.14)	0.12 (0.12)
Employment	-0.01 (0.12)	0.01 (0.11)	-0.01 (0.117)

Significance levels: *** 1%, ** 5%, * 10%. Standard errors reported in parenthesis.
Estimates are from a linear probability model, x100.
Estimation of (10) of the year 2003 and setting the placebo reform in February, May and November of that year.
Estimation through individual fixed effects. S.e. clustered at the individual level.

5. Conclusion

This thesis 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. The findings suggest that there is a sizable gap in between these. This is consistent across groups of age and health conditions. I find evidence that a sizable fraction of this gap can be explained by (short term) labor market frictions. However, there may be other reasons. The unwillingness of treated claimants to work may be one of them. Another reason may be the rejection from DI of truly disabled individuals. If this is the case, the reform may have caused welfare losses among reassessed claimants.

The reform caused the highest exit from DI among claimants with musculoskeletal disorders, followed by these with mental, endocrine and cardiovascular diseases. The rise in their employment relative to the drop in DI does not follow this same order. It seems that those claimants suffering from mental disorders have the hardest time in becoming employed. Again, it could be a sign of their inability or unwillingness to work, or the impossibility to be hired. The relative employment response remains constant across groups of age, with the exception for those claimants suffering from musculoskeletal disorders. I find that the ratio between their DI ouflow rise and employment increase is larger for younger claimants, suggesting that they move from DI to employment more easily than their older peers. This could be interpreted as the interaction between age and musculoskeletal disorders causing a larger impediment for work.

The focus of this thesis has been on transitions to employment, rather than actual earnings rebound. It could well be that despite 50 to 60 percent of the DI outflow being accompanied by labor inflows, the proportion of earnings compensated is below this level. This can be the case if the stigma from having participated in the DI program entails earnings losses while returning to the labor market, or if the separation from the labor market generates a skill depreciation of the workers (Moore, 2014).

Some additional points are left unanswered. For instance, I find that there seems not to be an increase in the frequency of employment while remaining in DI. This, however, could be an artifact of the measure of employment constructed. Indeed, claimants remaining in DI and employed after reassessment may have increased their labor supply at the intensive margin. It could be interesting for future research to explore the number of days worked.

References

Autor, David H., and Mark G. Duggan. 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.

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.

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

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. (2014), 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.

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.

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

Table 11: Treatment Effects per Group of Age and Main Diagnosis

	Endocrine		Cardio.		Musc.		Psych.	
Aggregate								
DI Only	-2.05***	(0.34)	-1.77*	(0.92)	-5.53***	(0.29)	-3.52***	(0.24)
Employment and DI	-0.93***	(0.33)	-0.35	(0.9)	-2.06***	(0.29)	-0.88***	(0.22)

Significance levels: *** 1%, ** 5%, * 10%. Standard errors reported in parenthesis.
Estimates are from a linear probability model, x100.
Estimation through individual fixed effects. S.e. clustered at the individual level.

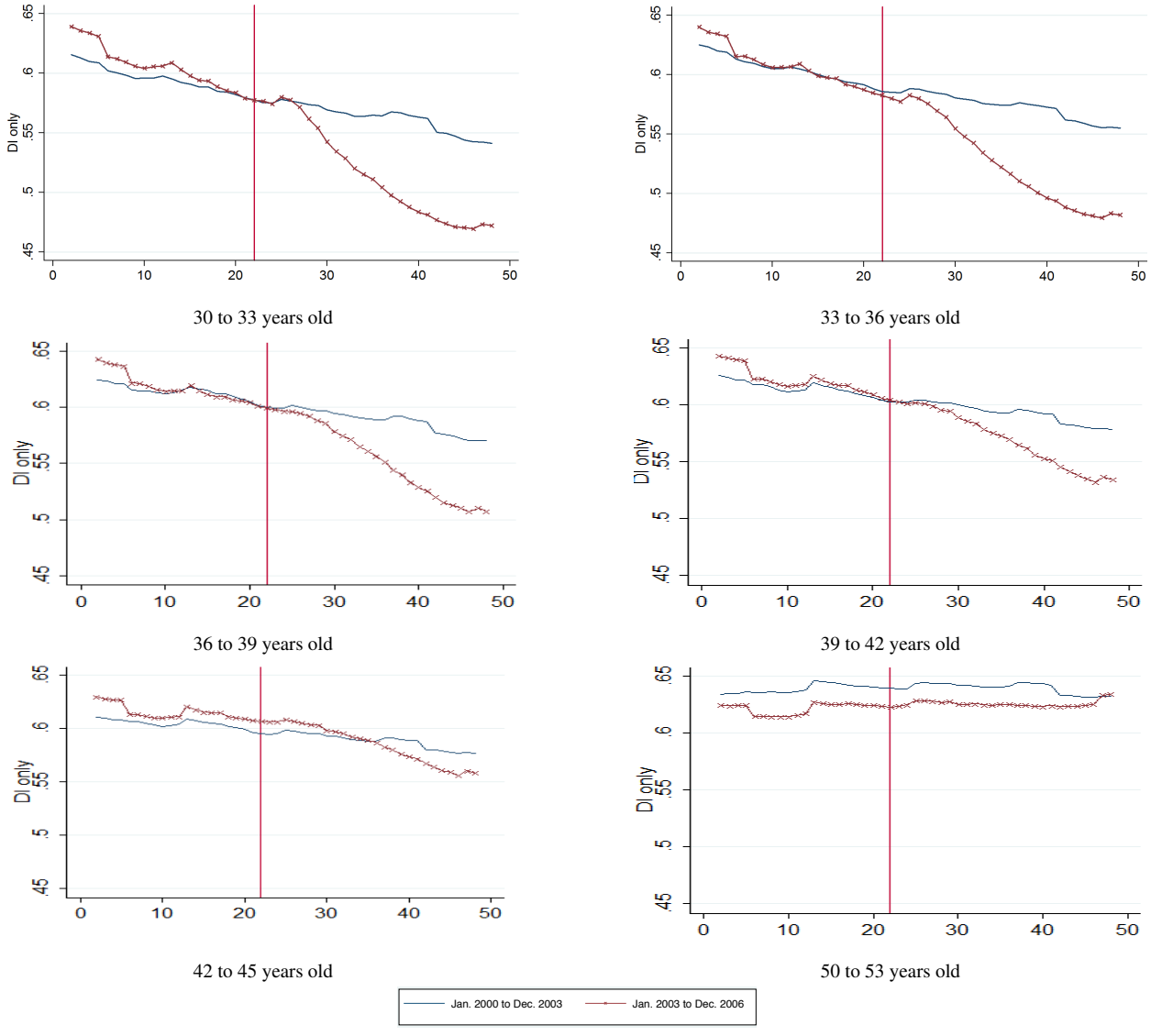


Figure 5: Probability to remain in DI Only

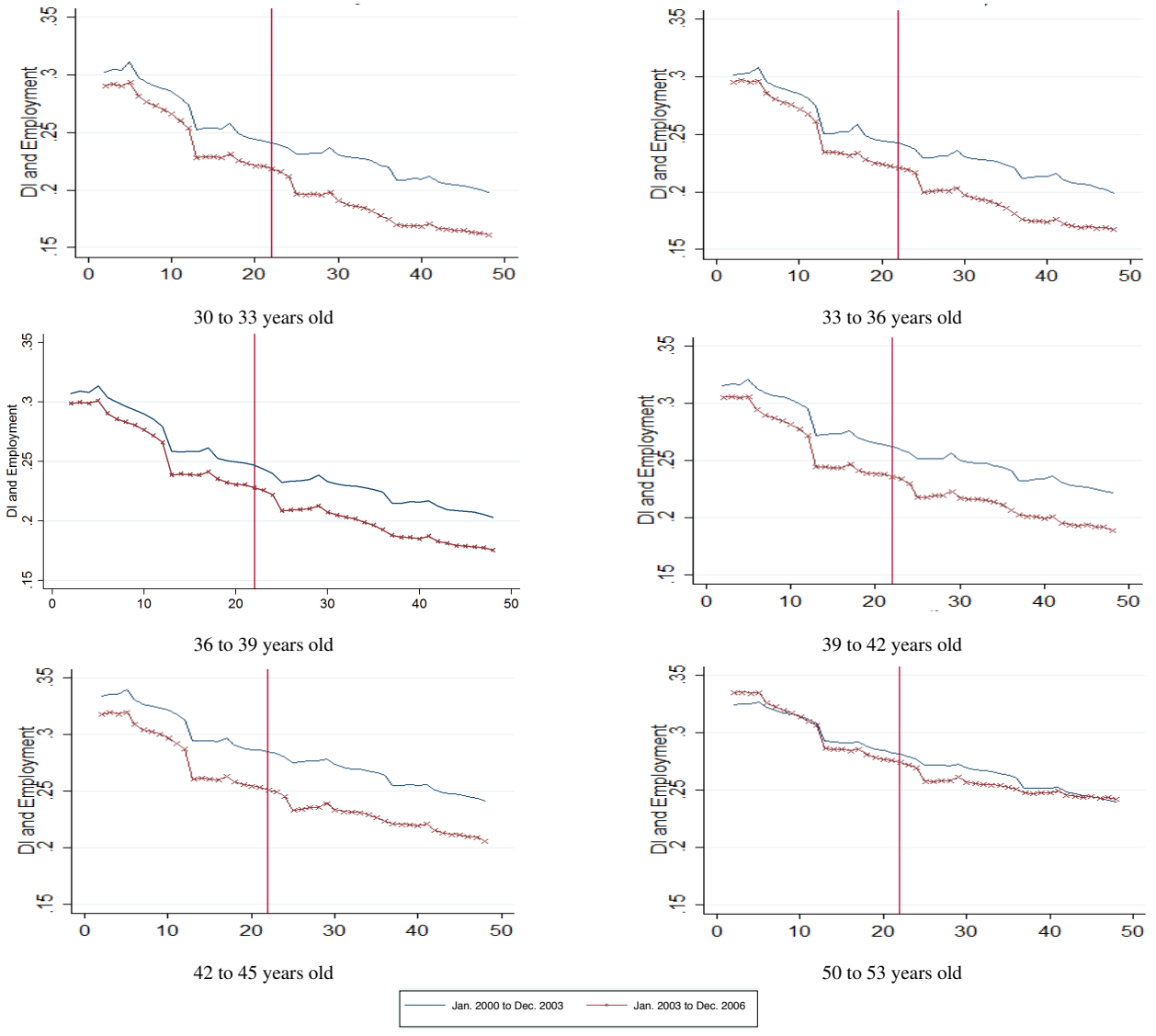
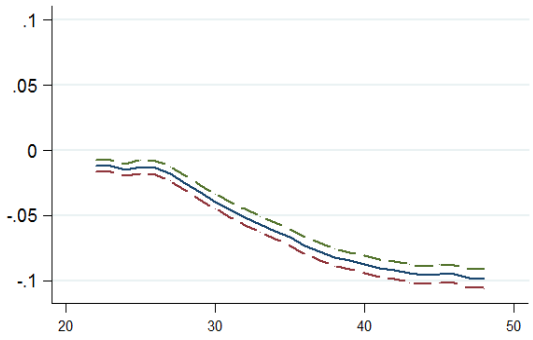
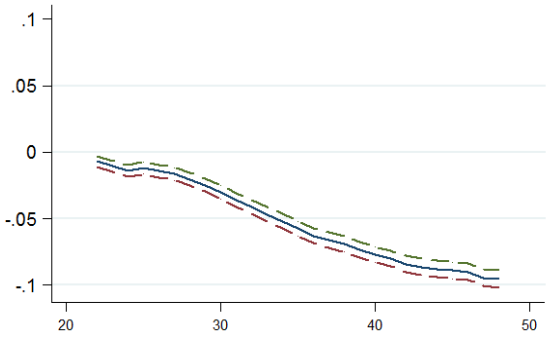


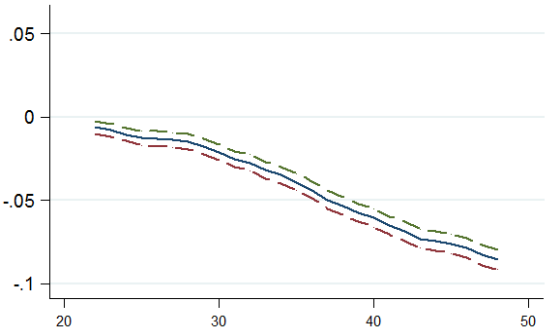
Figure 6: Probability to be Employed Only



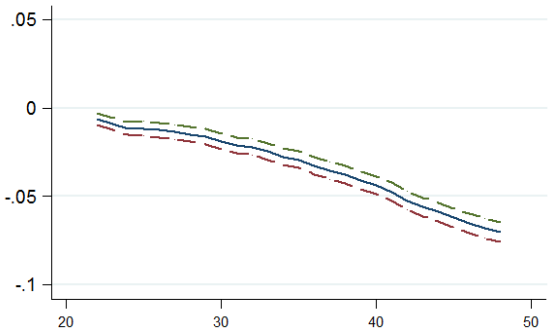
30 to 33 years old



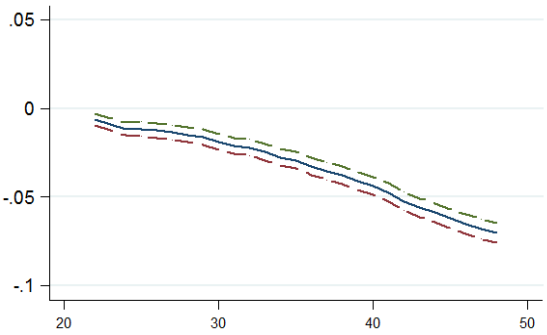
33 to 36 years old



36 to 39 years old

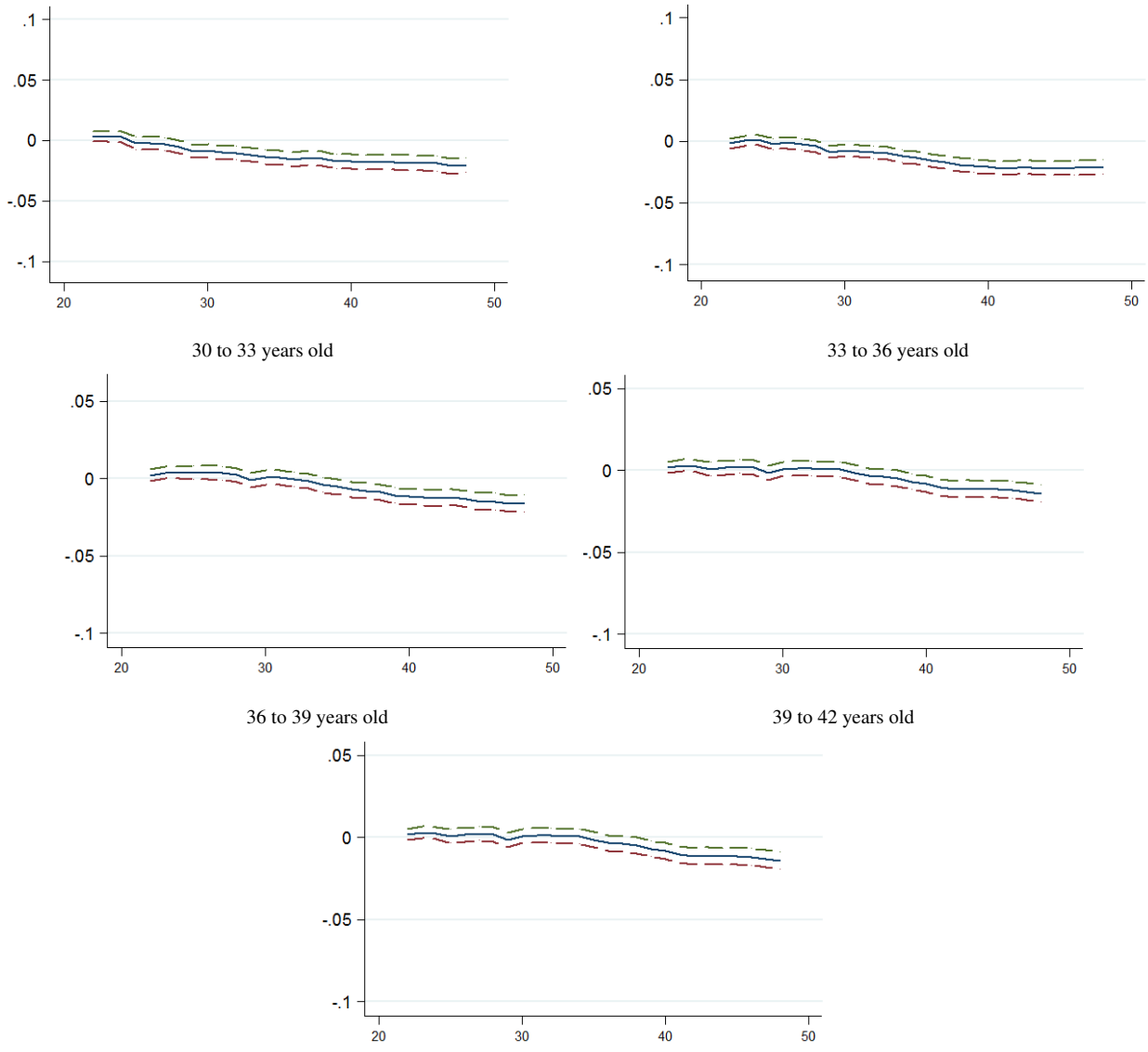


39 to 42 years old



42 to 45 years old

Figure 7: Flexible Specification: DI Only



30 to 33 years old

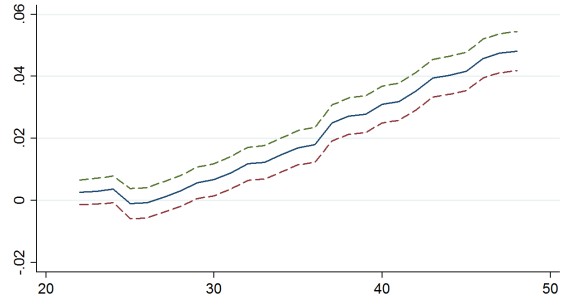
33 to 36 years old

36 to 39 years old

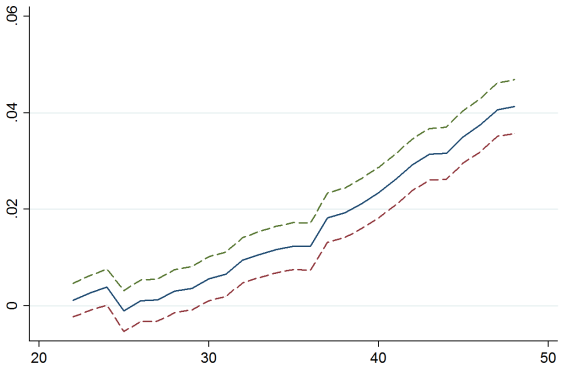
39 to 42 years old

42 to 45 years old

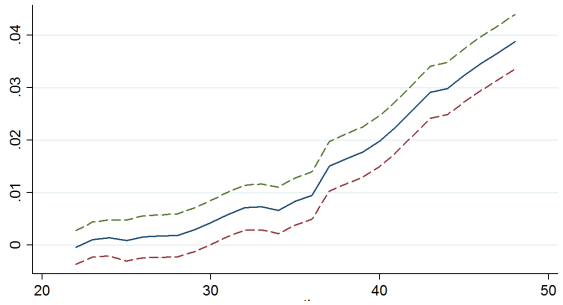
Figure 8: Flexible Specification Estimates: Probability to be employed and in DI



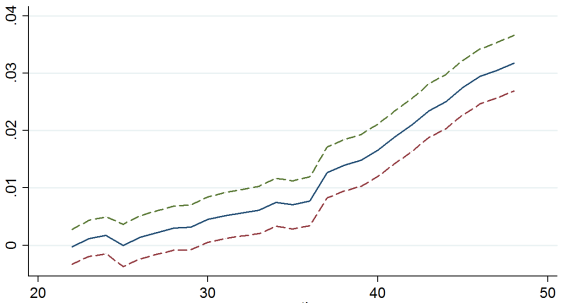
30 to 33 years old



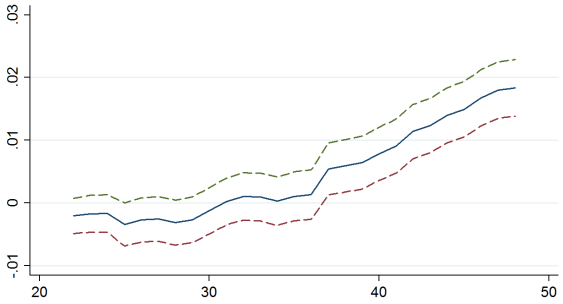
33 to 36 years old



36 to 39 years old



39 to 42 years old



42 to 45 years old

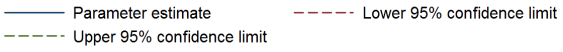


Figure 9: Flexible Specification Estimates: Probability to be Employed TOTAL