Earnings capacity of disability insurance recipients*

Sílvia Garcia-Mandicó, Pilar García-Gómez,
Anne C. Gielen, Owen O'Donnell†

December 2, 2018

Abstract

Determining the earnings capacity of benefit recipients is critical to the operation and evaluation of disability insurance (DI) programs. We use administrative data on the universe of Dutch DI recipients and reassessment of their benefit entitlement under more stringent rules to estimate earnings capacity they could be induced to utilize. We estimate that the increased stringency removed 17 percent from the program and reduced the amount of benefit received by 20 percent, on average. In response, employment increased by 6.7 percentage points and earnings rose by 18 percent. Recipients were able to increase earnings by €636 for each €1000 of DI income lost. This earnings response was largest from those with more subjectively defined disabilities (mental health and musculoskeletal conditions), as well as female and younger recipients. Consistent with earnings capacity deteriorating with claim duration, reassessment had the least impact on the benefits and earnings of those who had been claiming DI for longest. Working while claiming partial disability benefits may slow the deterioration of earnings capacity.

Keywords: Disability Insurance, Health, Employment, Earnings
JEL Codes: H53, H55, J14, J22

*We thank Dinand Webbink, Pierre Koning, Gabriele Ciminelli, Songul Togan, Coen Van de Kraats, Eric French, Gaute Torsvik, Andreas Kostol, Magne Mogstad, Nicole Maestas, Josef Zweimüller, and participants at various seminars and conferences for valuable feedback.

†Garcia Mandicó: Erasmus University Rotterdam and Tinbergen Institute, garciamandico@ese.eur.nl; García-Gómez: Erasmus University Rotterdam and Tinbergen Institute, garciagomez@ese.eur.nl; Gielen: Erasmus University Rotterdam, Tinbergen Institute and IZA, gielen@ese.eur.nl; O'Donnell: Erasmus University Rotterdam, Tinbergen Institute and University of Lausanne, odonnell@ese.eur.nl

1 Introduction

Disability insurance (DI) is intended to compensate for lost earnings capacity. The difficulty lies in determining how much has been lost. Overly stringent assessment leaves people underinsured. Overly lax assessment encourages moral hazard. Establishing whether the right balance has been struck requires finding a way to estimate the earnings potential of benefit recipients. A popular approach is to use the earnings of rejected DI applicants as the counterfactual (Bound 1989). This does not allow for the possibility that earnings capacity deteriorates while the benefit is being claimed. Consequently, it may produce upwardly biased estimates of the earnings potential of benefit recipients and their earnings response to any reduction in benefits.

This paper provides evidence on earnings capacity in the universe of Dutch DI recipients. It uses administrative data to estimate the impact on earnings, and employment, of a 2004 reform that involved reassessment of the earnings capacity of the entire stock of younger DI recipients under more stringent criteria, resulting in termination of some claims and substantial cuts in benefits. If the reassessments were effective in identifying recipients with residual earnings capacity, then the reductions in benefits should have induced increases in earnings. If, on the other hand, the reassessments were overly aggressive or poorly targeted, then the earnings response would be muted. We estimate the average effect of reassessment on earnings and scale this by the average reduction in benefits. The scaled effect is informative of the extent to which, on average, the upward revision of earnings capacity was justified. Heterogeneity in the reassessment of earnings capacity is explored by comparing benefit reductions across groups distinguished by length of claim, diagnosis, degree of disability, age and sex. Targeting is assessed by examining variation across these groups in the earnings response relative to the reduction in benefits.

We identify the effect of the reform by comparing the change in earnings (and other outcomes) of DI recipients aged 30-44, whose entitlement was reassessed under stricter criteria, with the respective change among older recipients, who were not reassessed. Unlike

studies that rely on difference-in-differences (DID) across age groups to identify effects of more stringent criteria at *application* for DI (Karlström et al. 2008; Staubli 2011), we adjust for the difference between the age groups in the outcome trend over a period prior to the reform. This trend-adjusted DID (Bell et al. 1999) eliminates age-specific trends, as well as period effects. Identification rests on the assumption that, in the absence of the reform, the age differential in the outcome trends would have been that observed in the earlier period. Consistent with this, we demonstrate that the age differential in the trends is similar over multiple periods prior to the reform. A placebo test also lends credibility to the identification: implementing the empirical strategy with data on individuals who are not DI recipients, we find no "effect" of a pseudo reform on earnings.

We estimate that, on average, application of the more stringent rules reduced the amount of DI income received by 20 percent, raised employment by 20 percent and increased earnings by 18 percent; implying high elasticities of employment and earnings with respect to benefits. A €1000 reduction in disability benefits is estimated to have been replaced by €636 of labor earnings. Apparently, some Dutch DI recipients had considerable untapped earnings capacity that benefit cuts arising from reassessment induced them to utilize.

The Netherlands provides an interesting context in which to assess the earnings capacity of DI recipients. It is known for a high DI dependency rate that reached 12% of the insured population at the beginning of the 1990s, while also being commended for a series of reforms, such as the one we examine, that are claimed to have contributed to a two-fifths reduction in this dependency (Koning and Lindeboom 2015). Officials and analysts in countries, such as the US, looking for ways to manage the escalating fiscal burden of DI can potentially learn from the Dutch experience (Autor 2015; Burkhauser et al. 2014). By examining a reform that occurred a decade into the rationalization of an initially generous program, when the dependency rate had already fallen considerably (Koning and Lindeboom 2015), we deliver evidence that is more relevant to the situation prevailing in other countries than the evaluation by Borghans et al. (2014) of an earlier Dutch reform that took effect when

DI enrollment was substantially higher than elsewhere. Our estimate of the rate at which earnings replaced lost DI income is actually very close to that obtained by Borghans et al. (2014), indicating that even after a decade of retrenchment some DI recipients still had considerable unused earnings capacity they could call on to replace around two thirds of substantial cuts to benefits. However, it is important to emphasize that these were the minority of the stock of DI recipients. Most did not have their benefits reduced despite being subjected to reassessment under more stringent criteria.[1]

Much of the evidence on the earnings crowd-out from DI comes from studies that follow Bound (1989) in using the earnings of rejected applicants to place an upper bound on the earnings capacity of successful applicants (Chen and Van der Klaauw 2008; Von Wachter et al. 2011). Exploitation of plausibly exogenous variation in the award or appeal probability can eliminate upward bias in estimated earnings capacity at the time of application (Autor et al. 2017; French and Song 2014; Maestas et al. 2013). However, this strategy will underestimate the impact of DI on employment and earnings if, as Autor et al. (2015) demonstrate for the US Social Security Disability Insurance (SSDI) program, work capacity, preferences or both decay during the time it takes for an application to be adjudicated. Further, it will overestimate the earnings potential of the stock of beneficiaries if skills and preferences for work deteriorate with time spent on DI. Evidence obtained from comparison of accepted and rejected applicants is pertinent to the impact of policies that tighten entry to DI. It is less relevant to assessing the potential of reforms, such as the one we examine, that aim to release the residual earnings capacity of benefit recipients.

Besides being one of the few studies to estimate earnings responses to targeted reductions in the benefit entitlement of DI recipients, this paper makes three main contributions to evidence on the earnings capacity of claimants. First, it adds to the meager evidence on

[1]Our estimates of high elasticities of employment and earnings with respect to DI benefits may appear to contradict the claim of Koning and Lindeboom (2015) that benefit reductions played a minor role in lowering DI dependency in the Netherlands. However, their argument is that tightening of eligibility criteria has contributed less to reductions in *entry* to DI than measures that operated through employers' obligations and incentives. They acknowledge the large effect on exit of the 1993 reform evaluated by Borghans et al. (2014) but do not mention the 2004 reform we evaluate.

whether and how earnings capacity varies with time spent on DI. The longer someone is claiming, the more their skills may be expected to deteriorate and their work preferences dissipate (Bryngelson 2009; Svensson et al. 2010; Vingård et al. 2004). These negative effects of claim duration are potentially offset by time on DI providing the opportunity for health, and so earnings capacity, to partially recover from any illness that does not cause permanent disability. Moore (2015) finds that, in the US, the employment response to DI termination increases with claim duration up to a spell of about three years, but thereafter the relationship turns negative. This is consistent with the health recovery effect first dominating before being overcome by the labor market detachment effect. But there is little or no supporting evidence for this inverted U-shaped relationship between work capacity and DI duration. Autor et al. (2015) find that employment and earnings fall even as the time waiting for a US SSDI application to be decided lengthens. Gelber et al. (2017) estimate income effects of SSDI benefits on earnings and find that they vary little with the duration of a claim. Using claim durations of up to 15 years, which is substantially longer than in these other studies, we also do not find evidence of an inverted U relationship. Reassessment of earnings capacity under less generous rules causes those who had been claiming DI for longer to increase their earnings by less.

Second, the paper assesses whether the option of partial disability helps DI recipients retain their earnings capacity. The Dutch program's provision for partial disability, which permits beneficiaries to permanently earn in excess of the equivalent of the substantial gainful activity (SGA) limit beyond which SSDI benefits are terminated in the US, makes it a forerunner of return-to-work incentives introduced or contemplated elsewhere (Kostol and Mogstad 2014). Autor (2015) argues that SSDI should introduce the option of partial disability to reduce the tendency for benefit recipients to become detached from the labor market. If the measure achieves this, then the earnings capacity of the partially disabled who continue to work should decline less with time on DI. Our estimates, which are obtained from a sample in which approximately half of the claimants targeted by the reform are partially

disabled, are consistent with this hypothesis. The positive earnings impact of the increased stringency declines steeply with claim duration for those who were not working at the time of reassessment. In contrast, among the partially disabled who were working, the effect on the earnings of those who had been on DI for 15 years is as strong as that after a claim spell of only a year.

The strong, positive effect on the earnings of the partially disabled contrasts with Koning and van Sonsbeek (2017), who find, also using Dutch administrative data, that a financial penalty imposed on partially disabled claimants who do not utilize their residual earnings capacity has only a small effect on their earnings, although it does raise part-time (but not full-time) employment. The difference in the magnitudes of the earnings responses may be because Koning and van Sonsbeek (2017) estimate the effect of a non-targeted benefit reduction the recipients could anticipate, while we estimate the response to unanticipated reductions targeted on those assessed to have greatest earnings capacity.

Third, because the qualification rules were tightened for all recipients irrespective of the nature of their disability, we are able to compare earnings capacity across DI beneficiaries qualifying through different diagnoses. This is relevant to assessing the credibility of one of the main hypotheses advanced to explain lengthening DI rolls. That is, loosening of the criteria for entitlement from precisely defined medical diagnoses to the vaguer concept of *work capacity* opened the door to claims based on health conditions that are difficult to verify medically (Autor 2015; Autor and Duggan 2006).[2] If this explanation for the growth of DI programs is correct, then claimants with more subjectively defined conditions should have greater earnings capacity. Comparisons between accepted and rejected DI applicants in the US produce contradictory evidence on the validity of this hypothesis (French and Song 2014; Maestas et al. 2013; Von Wachter et al. 2011).[3] Consistent with it, Moore (2015) finds that

[2]Lower back pain and stress-related problems are the stereotypical examples. In 2012 across all OECD countries, individuals suffering from mental health disorders constituted one half of DI benefit recipients and 27-47% of new awards (OECD 2012). Musculoskeletal problems are typically the second most common reason given for a DI claim.

[3]Von Wachter et al. (2011) find in favor of it. Maestas et al. (2013) correct for unobservable differences between accepted and rejected DI applicants and find no evidence that those reporting mental health and

US DI recipients with a primary diagnosis of a mental health or a musculoskeletal condition were more likely to work after their benefits were terminated. But this evidence is obtained using only the 2% benefit recipients who had partly qualified through an alcohol or drug addiction. The earnings capacity and preferences of this group would not be expected to be representative of those of other DI recipients. We provide the first test of the hypothesis using data on a universe of those receiving DI benefits, who are all subjected to the same tightening of entitlement rules that did not involve any change in the medical criteria for qualification. In both absolute and relative terms, benefits were cut more aggressively, reflecting greater upward revision to assessed earnings capacity, among recipients with mental health and musculoskeletal conditions compared with those qualifying through all other diagnoses. And those in the first two diagnostic groups, particularly those with a mental health condition, were able to replace a larger fraction of their greater loss of benefits with increased earnings. Hence, both the assessed and realized earnings capacity were higher for recipients with these more subjectively defined conditions.

The paper proceeds as follows. Section 2 outlines key features of the Dutch DI program and the reform we evaluate. Section 3 sets out our identification strategy. Section 4 describes the data and examines trends in the outcomes. Section 5 presents the results starting with full sample estimates, then a placebo test and robustness analysis, followed by examination of relationship between earnings capacity and duration of a DI claim before heterogeneous effects by type and degree of disability, age and sex are examined. The final section concludes.

musculoskeletal conditions have greater work capacity. In fact, they find that those with the latter type of condition have lower than average work capacity. French and Song (2014), who also correct for unobservables, arrive at the opposite conclusion: those with musculoskeletal problems have greater work capacity.

2 Disability insurance in the Netherlands

2.1 Eligibility and benefits

The 2004 reform changed the details but not the general procedures for assessing DI eligibility and benefit entitlement. Before describing the reform, we summarize those procedures.

An application for full disability benefits can be submitted after a period of sick pay, which was one year in 2004. Application for partial disability benefits can be made while in work. The Social Insurance Benefits Agency (UWV) conducts a medical assessment to establish whether the applicant is completely incapable of work. If the agency's physician judges that the applicant has some residual work capacity, then a vocational expert identifies specific occupations the applicant is considered capable of performing, taking educational attainment into account. Earnings capacity is then approximated by the average salary across the three highest paying of those occupations. *Degree of disability* is defined as the proportionate shortfall of this earnings capacity from pre-disability earnings. If this is below a threshold, which in 2004 was 15%, then the claim is rejected.[4] If it is at least 80%, then the applicant is classified as fully disabled and maximum benefits are paid. The claimant is compensated, at least initially, for approximately 70% of lost earnings capacity.[5]

The benefit recipient is permitted to do paid work without the loss of benefits but only up to the maximum earnings consistent with their assessed degree of disability. Earning

[4]The threshold was increased to 35% in 2006 for new applicants. This change did not affect the DI recipients we examine, who had all applied and were receiving DI before 2006. Neither did it affect reassessments of the entitlement of these recipients conducted after 2006.

[5]Specifically, the replacement rate is set at 70% of the mid-point of each interval of the degree of disability. The intervals are: [15%, 25%), [25%, 35%), [35%, 45%), [45%, 55%), [55%, 65%), [65%, 80%) and [80%, 100%]. The replacement rate in the top interval is 70%. Claimants who are less than fully disabled receive this earnings-related benefit for a limited period of up to 6 years that depends on their age and work history. After that, benefits remain constant and tied to pre-disability earnings only if the recipient exercises at least 50 percent of their assessed earnings capacity. Otherwise, their benefit income drops to a function of the minimum wage, plus a supplement based on age. However, at the time of the reform we study, about 80% of employees had supplementary insurance arranged through collective labor agreements that compensated for the drop in social insurance payments after the initial period (Van Vuren and Van Vuren 2007). Using variation in the length of this earnings-related benefit period, Koning and van Sonsbeek (2017) estimate that the change in work incentives at the transition raises employment of the partially disabled by 2.5 percentage points but, as mentioned in the introduction, has little impact on earnings and no effect on full-time employment.

more than that results in downward revision of the degree of disability and a reduced benefit payment. After leaving DI, benefits continue to be received during a three-month trial period before entitlement is lost. Prior to the reform, outflow from DI was low. The degree of disability was reassessed one year after a claim was awarded and every five years thereafter. These reassessments were often based on no more than the recipient's response to a postal questionnaire.

2.2 The Reform

From October 2004, the stock of DI benefit recipients younger than 50 on July 1, 2004 became eligible for reassessment under more stringent criteria.[6] Reassessment had two components. First, recipients were required to undergo a medical examination that could result in a revision to their assessed functional limitations. Descriptive analysis presented in Appendix A suggests that this stage contributed rather substantially to reducing benefit entitlements. Second, the degree of disability was re-calculated using stricter rules that could result in upward revision of earnings capacity and downward revision of pre-disability earnings. This part of the reform intensified the reduction in entitlement through downward revision of the degree of disability that began with the 1993 reform evaluated by Borghans et al. (2014).[7] As a result, for any given health condition and associated functional limitations, the degree of disability would either be reduced or remain unchanged. Consequently, the benefit paid

[6]The reform was legislated in April 2004 and the initial plan was to start the reassessments from July 2004. Political opposition and lack of consensus about the reassessment criteria resulted in implementation being pushed back to October. Analysis in section 4.3 of trends in employment and earnings prior to the start of the reassessments does not reveal patterns consistent with anticipation effects.

[7]Earnings capacity was now estimated by averaging over the three highest paying occupations considered compatible with the recipient's functional limitations in which there were at least three active workers engaged in the local labor market. Previously, the averaging was done over the three highest paid occupations with at least ten workers. In addition, jobs requiring Dutch language proficiency and knowledge of information technology were now considered feasible even if the recipient did not possess those skills. Full-time employment and night work were now also considered feasible even if the recipient had not previously engaged in them. As a result of expanding the pool of potential work, the average earnings over the highest paying occupations within the pool, and so earnings capacity, was likely to rise. It could not fall. A rise in assessed earnings capacity translated into a fall in benefit entitlement. Further, pre-disability earnings could be reduced by a new rule that truncated weekly hours at a maximum of 38. If earnings had been inflated by previously working more than this, then there would be a downward revision of lost earnings capacity, and so benefits.

could be cut or terminated.

The reassessments were undertaken between October 2004 and April 2009. However, very few (1.2%) were done in 2004, almost half (46%) were performed by the end of 2005 and they were all but completed (99.9%) by the end of 2008 (Table 1). Initially, the plan was to reassess all younger benefit recipients before moving to older groups, but this was not observed. The order in which recipients were called for reassessment was, however, far from random. It is correlated to the outcome of the reassessment (see Appendix A). It appears that the benefits agency was more likely to call for earlier reassessment those recipients who were anticipated to have greater surplus of current over previously assessed earnings capacity, and so would experience larger cuts to benefit. For this reason, we do not attempt to exploit variation in the timing of reassessment within the treatment group for identification.

Table 1: Timing of reassessments of DI recipients aged 30-44 on July 1, 2004

Year	Frequency	Cumulative %
2004 (Oct-Dec)	1,642	1.2
2005	61,724	46.0
2006	48,197	81.0
2007	21,459	96.5
2008	4,593	99.9
2009	199	100.0
TOTAL	137,814	

Source: Authors' calculations using data provided by the Social Insurance Benefits Agency (UWV). The sample is selected as our treatment group, i.e. the stock of DI recipients in January 2004 who were aged 30-44 on July 1, 2004. However, it excludes those who exited DI before being called for reassessment.

In 2007, strong criticism of the policy and a change of government resulted in the age threshold for reassessment being revised from less than 50 to less than 45 on July 1, 2004. As a result, around 17,000 recipients aged 45-49 who had already been reassessed were assessed once more under the old, more lenient rules (Ministry of Justice 2007).[8] Consequently, we

[8]Those aged 45-49 who were reassessed twice under different rules appear to be exceptional in the extent to which their degree of disability was initially reduced (see Appendix A). This probably reflects targeting for earlier reassessment those recipients who were expected to be most affected by it. It rules out using differences in exposure to reassessment within this age group for identification.

restrict attention to benefit recipients aged 30-44 on July 1, 2004.

The transition matrix presented in Table 2 is constructed from DI recipients aged 30-44 at the time of the reform. It shows, for each degree of disability interval before reassessment, the percentage of recipients in each interval after reassessment. About a quarter (24.4%) were reassessed as having a degree of disability below the 15% minimum threshold and had their entitlement withdrawn completely.[9] The fraction of those initially with the lowest degree of disability [15%, 25%) who were disqualified is almost one half. Even among those who initially were classified as fully disabled ([80%, 100%] interval), 17% were placed below the minimum threshold after reassessment and lost their benefits entirely. About 10% of recipients were allowed to remain on DI but with lower benefits. Consequently, more than a third (34.4%) had their benefits either cut or terminated. A majority (58.5%) experienced no change in their entitlement.[10] The initially fully disabled were least affected: 72% continued to received the same amount of benefit.[11] Despite the application of more stringent rules, 6% of recipients had their degree of disability raised following reassessment because the medical reexamination detected a deterioration in health and increased functional impairment since the previous assessment.

The consequences of the reform for benefit entitlement were clearly heterogeneous. Greater upward revision to assessed earnings capacity, and so downward revision to degree of disability, resulted in larger reductions in benefits. We are not estimating the effects of an across-the-board benefit cut. Rather, we estimate the average effect of reassessment on benefit income, as well as the average effects on employment and earnings resulting from the targeted revisions to benefit entitlement. These effects are obtained by averaging over all who were reassessed, a majority of whom experienced no change in their benefit entitlement. The average effect on benefit income will be much smaller than the average reduction in

[9]First column total as a percentage of the overall total of Table 2.

[10]The sum of the frequencies used to obtain the cell entries on the diagonal of the table as a percentage of the overall total is 58.5%.

[11]This group included some who were not called for medical examination because their full disability was apparent from the seriousness of their condition identified on file. These case files were reviewed, however. The reform involved reassessment of the degree of disability of all benefit recipients aged 30-44.

benefits paid to those whose earnings capacities were raised as a result of reassessment.

Table 2: Degree of disability before and after reassessment transition matrix, recipients aged 30-44 on July 1, 2004

	After (%)								
	<15%	[15,25)	[25,35)	[35,45)	[45,55)	[55,65)	[65,80)	[80,100]	N
Before									
[15,25)	47.9	35.5	6.3	1.8	0.9	0.8	0.4	6.4	12,498
[25,35)	29.7	18.7	34.9	5.6	2.0	1.0	0.8	7.4	11,554
[35,45)	23.8	9.8	16.4	31.8	5.8	1.6	1.4	9.5	8,210
[45,55)	25.8	5.3	6.4	9.6	32.3	4.2	2.1	14.3	7,562
[55,65)	23.7	6.9	5.6	5.6	10.6	25.9	4.6	17.1	4,281
[65,80)	16.6	5.5	6.9	7.2	7.1	8.4	26.7	21.6	3,574
[80,100]	17.1	2.5	2.1	1.8	1.9	1.4	1.5	71.9	90,135
N	33,585	11,069	8,661	5,842	5,274	3,102	2,594	67,687	137,814

Notes: Each row shows, for a given degree of disability before reassessment, the percentage of recipients at each degree of disability after reassessment under more stringent rules. Row percentages sum to 100. Authors' calculations using data provided by the Social Insurance Benefits Agency (UWV). Sample is defined as explained in the notes to Table 1.

If the outcome of reassessment was a downward revision in the degree of disability, then benefits were reduced or terminated two months later. If employment was not secured, a disqualified DI recipient could transfer to unemployment insurance (UI) if still eligible for that program. If not, or if UI entitlement would last for less than six months, then application could be made to a temporary program put in place specifically to cushion the short term impact of the reform. This maintained DI income at the same level for a period of six months (increased to twelve months in 2007). Around 18% of recipients whose entitlements were reduced or terminated were granted benefits from this program (Social Insurance Benefits Agency (UWV) 2009).

Further details of the implementation of the reform and the reassessment process are given in Appendix A.

3 Identification & Estimation

3.1 Identification

We estimate effects on benefit receipt and labor supply of reassessment, which consists of medical examination to reconsider functional impairment and application of more stringent rules to calculate degree of disability (and so benefit entitlement) given functional impairment. To estimate the average effect of reassessment on recipients aged 30-44, we need a comparison group(s) that allows credible identification of the average outcome that would have materialized in the target group if the reform had not been implemented.

Let Y_{it} be the observed outcome of individual i at time t, and let Y_{it}^1 and Y_{it}^0 represent potential outcomes with and without reassessment respectively. Let *t=0* indicate some time before the commencement of reassessments, such that $Y_{i0} = Y_{i0}^0 \ \ \forall i$. In our main analysis, we use annual data and *t=0* corresponds to 2004. This introduces a slight inaccuracy since around 1% of recipients aged 30-44 were reassessed in the last quarter of 2004 (Table 1). We test robustness to using monthly data, which avoids this inaccuracy, in section 5.3.[12] Let *t=4* be four years later in 2008 when the reassessments were completed (but for a negligible $< 0.001\%$, Table 1). Then, $Y_{i4} = D_i Y_{i4}^1 + (1 - D_i) Y_{i4}^0$, where $D_i = 1$ if i has been reassessed and is 0 otherwise. We wish to estimate the average effect of reassessment on those reassessed: $ATET = \mathbb{E}\left[Y_{i4}^1 - Y_{i4}^0 \mid D_i = 1\right]$.

One potential identification strategy would rely on a difference-in-differences (DID) comparison between younger benefit recipients (30-44 on July 1, 2004) who were subject to reassessment and older recipients (50+ on July 1, 2004) who were not.[13] This is likely to be problematic, particularly as the age gap widens. Older DI beneficiaries have a lower probability of returning to work and recovering their earnings than younger recipients, even

[12] We do not use monthly data throughout because they are more noisy and the dataset becomes extremely large, which slows computation considerably on the remote server through which the administrative files are accessed.

[13] Those aged 45-49 on July 1, 2004 are not useful either as a treatment group or a comparison group since some of them were first reassessed under the new, stricter rules and then (after 2007) assessed once again under the initial, more lenient rules.

when the latter are not subject to reassessment. An alternative comparison group would be DI recipients who are the same age as those targeted by the reform but who are observed in a period that ends before the reassessments begin. The threat to a DID strategy using this comparison group comes from period-specific labor market conditions and any earlier changes in DI that would invalidate using the earlier period to identify counterfactual employment and earnings of the target age group in the reform period.

Our strategy makes use of both comparison groups – older benefit recipients in the same period and recipients of the same age in an earlier period – to identify the impact of reassessment under an assumption that is plausibly (although not necessarily) weaker than each assumption required to construct the counterfactual from one of the two comparison groups alone. We use a four-year interval running from 1999 to 2003 ($PERIOD_i = 0$) that precedes the reform to identify the extent to which the trend in the average outcome of younger DI recipients aged 30-44 ($AGE_i = 1$) differs from the trend of older recipients, whom we define as aged from 50 to 53 ($AGE_i = 0$) for reasons given in section 3.2. Effectively, we subtract the age-differential trend in the *non-reform period* from the DID of the outcome between the age groups over the four-year *reform period* running from 2004 to 2008 ($PERIOD_i = 1$) during which the younger age group was reassessed. This differential trend adjusted difference-in-differences (DADID) (Bell et al. 1999; Blundell and Costa Dias 2002) relaxes the assumption of common trends in earnings (/employment) across age groups in the absence of the reform. It also avoids the assumption of common trends in earnings for a given age group across periods. The assumption that is required for identification of the $ATET$ by DADID is that the age differential in the trends in earnings would have been common across periods in the absence of the reform:

$$\begin{aligned} &\mathbb{E}\left[Y_{i4}^0 - Y_{i0}^0 \mid AGE_i = 1, PERIOD_i = 1\right] - \mathbb{E}\left[Y_{i4}^0 - Y_{i0}^0 \mid AGE_i = 0, PERIOD_i = 1\right] \\ =&\mathbb{E}\left[Y_{i4}^0 - Y_{i0}^0 \mid AGE_i = 1, PERIOD_i = 0\right] - \mathbb{E}\left[Y_{i4}^0 - Y_{i0}^0 \mid AGE_i = 0, PERIOD_i = 0\right] \end{aligned} \tag{1}$$

If this assumption holds, then any widening of the age differential in the trends that occurs in

the reform period relative to the non-reform period can be attributed to a positive impact of reassessment on the earnings of younger benefit recipients. The average effect of reassessment on those reassessed is then given by the DADID:

$$\begin{aligned}
&\mathbb{E}\left[Y_{i4} \mid AGE_i = 1, PERIOD_i = 1\right] - \mathbb{E}\left[Y_{i0} \mid AGE_i = 1, PERIOD_i = 1\right] \\
&- \Big(\mathbb{E}\left[Y_{i4} \mid AGE_i = 0, PERIOD_i = 1\right] - \mathbb{E}\left[Y_{i0} \mid AGE_i = 0, PERIOD_i = 1\right]\Big) \\
&- \Bigg\{ \Big(\mathbb{E}\left[Y_{i4} \mid AGE_i = 1, PERIOD_i = 0\right] - \mathbb{E}\left[Y_{i0} \mid AGE_i = 1, PERIOD_i = 0\right]\Big) \\
&- \Big(\mathbb{E}\left[Y_{i4} \mid AGE_i = 0, PERIOD_i = 0\right] - \mathbb{E}\left[Y_{i0} \mid AGE_i = 0, PERIOD_i = 0\right]\Big) \Bigg\}
\end{aligned} \tag{2}$$

In section 4.3, we assess the plausibility of the assumption that the age differential in the earnings (/employment) trends would have been the same across the periods if no reassessments had been conducted between 2004 and 2008 by comparing age differences in trends across periods in which there was no reform. If the reform was anticipated by benefit recipients who reacted by leaving DI and entering employment already in 2004, then our strategy will deliver lower bound estimates of the effect. But examination of pre-reform trends does not reveal patterns consistent with anticipation. In section 5.2, we further assess the credibility of the strategy by checking that it gives a zero "effect" on the earnings (/employment) of individuals who were not DI recipients and so were not exposed to the reform.

3.2 Estimation

To estimate the effects, we pool balanced panels of DI recipients from the reform and non-reform periods. At entry to the panel, which is January 1, 2004 and January 1, 1999 for the reform and non-reform periods respectively, every observation is receiving DI benefits. In the reform period panel, the treated recipients are aged 30-44 on July 1, 2004. The comparison group obtained from this panel is aged 50-53 on July 1, 2004. We choose this age range in order to obtain a comparison group that is sufficiently large while remaining reasonably close to the treatment group in age, which makes the identification assumption more credible. In

section 5.3, we demonstrate robustness to using narrower and wider age intervals to define the comparison group. In the non-reform period panel, we distinguish between those aged 30-44 and those aged 50-53 on July 1, 1999.

We use least squares to estimate fixed effects models with the following structure,

$$Y_{it} = \sum_{t=1}^{4} \Big(\beta_t AGE_i \times PERIOD_i \times YEAR_t + \theta_t YEAR_t + \gamma_t AGE_i \times YEAR_t + \delta_t PERIOD_i \times YEAR_t \Big) + \mu_i + \varepsilon_{it}, \tag{3}$$

where $YEAR_t$ is an indicator of the within panel year of the observation, such that $YEAR_0 = 1$ & $PERIOD_i = 1$ indicates 2004, $YEAR_0 = 1$ & $PERIOD_i = 0$ indicates 1999 and $YEAR_4 = 1$ indicates 2008 or 2003 depending on the value of $PERIOD_i$, μ_i is an individual fixed effect and ε_{it} is an idiosyncratic error. In addition to period effects and age effects that differ between the periods, both of which are captured by the fixed effects, this model allows within panel time effects (θ_t) that differ across age groups (γ_t) and periods (δ_t). The period-specific level effects and trends allow for the fact that the periods 1999-2003 and 2004-2008 span different phases of the business cycle. Growth was decelerating in the earlier period and accelerating in the later period. The age-specific trends allow for the possibility that, within each period, average earnings (employment) of the younger group of DI recipients does not move in parallel to that of the older group.

Subject to the identification assumption (1), β_t corresponds to the average effect of the reform t years after it started to be implemented. Prior to $t = 4$, corresponding to 2008 in the reform period, the effects are not so interesting since not all benefit recipients in the target group aged 30-44 had been reassessed before then (Table 1). We focus on the estimate of β_4, which corresponds to the $ATET$ of reassessment. Note that we are estimating the effect of reassessment, not of the reduction in benefits that is the consequence of some, but not all, reassessments. By estimating the effect on benefits received, as well as on earnings (and employment), we can assess the extent to which earnings capacity was revised upwards, and we can examine the responsiveness of earnings (employment) to reduced benefit entitlement.

We cannot estimate effects after 2008 because this would require extending the length of the non-reform period, which is impossible since data are not available before 1999 and the reform period begins after 2003.

4 Data

4.1 Sources and measures

We obtain data on all recipients of DI benefits from social security files, which record degree of disability, benefit amount, claim duration and main diagnosis. We use these data to estimate the effect of the reform on the probability of receiving DI and the (annual) amount received. Diagnosis recorded on entry to DI is used to distinguish claimants in the two diagnostic groups that include the most subjectively defined disabilities - musculoskeletal conditions and mental disorders. We lump all other disabilities together. The social security files are also used to identify benefits received from other social insurance and social assistance programs, which we aggregate to obtain annual net of tax income from social transfers other than DI.

Data on employment, days worked and annual earnings (net of tax) are taken from files (polisadministratie) maintained by the Social Insurance Benefits Agency (UWV) that contain information related to income sources subject to earnings tax. We count a person as employed if registered as an employee for at least one day in a calendar year.

Municipal registers are used to identify date of birth and gender. Deaths are identified from the mortality register. The administrative files are linked using a unique individual identification number (RIN-code) that is issued on compulsory registration with the municipality at birth or after immigration. Additional details of the data sources and measures are provided in Appendix B Table B1.

4.2 Treatment and comparison groups

To construct the reform period sample, we select individuals who were claiming DI in January 2004. Of these, 3.9% died before the end of 2008 and are dropped from the panel. Mortality obviously differs between the age groups. But the age differential in mortality rates does not differ between the reform and non-reform periods. Hence, conditioning on survival does not introduce any compositional change that would bias the DADID estimates. We drop benefit recipients aged 45-49 on July 1, 2004 because of their inconsistent exposure to the reform that we described above. We also exclude recipients younger than 30 because there are very few of them and they typically have had little employment experience. Their employment patterns are likely to differ markedly from the older claimants we use as one comparison group. This leaves a treatment group of 160,194 individuals who were claiming DI in January 2004, were aged 30-44 on July 1, 2004 and so were eligible for reassessment and could be followed to the end of 2008 when the reassessments were completed.[14]

One of our comparison groups comprises 94,404 individuals who were claiming DI in January 2004, were aged 50-53 on July 1, 2004 and so were not subject to reassessment. The non-reform period sample consists of individuals who were claiming DI in January 1999, were aged either 30-44 (as the treatment group, 139,524 individuals) or 50-53 (as reform period comparison group, 102,464 individuals) on July 1, 1999, and survived to the end of 2003. We pool this balanced panel with that constructed for the reform period.

Table 3 shows means of characteristics at selection into the samples, i.e. 1999 and 2004, by age group and period. In both age groups, there is a higher fraction of females in the later period. This partly reflects increasing labor force participation of Dutch women and is consistent with the feminization of DI rolls observed in other countries. More relevant to the plausibility of our identification strategy is that the age group difference in the proportion

[14]The number of benefit recipients from this group who were reassessed is 137,814. The others left DI before being called for reassessment. Since this exit may have been in anticipation of the outcome of reassessment, these individuals can be considered to have been exposed to the reform and are rightly part of the treatment group. Their inclusion will downwardly bias the DADID estimate only if they exited already in 2004 (or October 2004 with the monthly data used in robustness analysis).

of female benefit recipients is roughly constant across the two periods. The same is true with respect to the average duration of a DI claim and the amount received. There is a discernible age group difference in the proportion of fully disabled claimants only in the earlier, non-reform period. Related to this, only in this period does the employment rate differ across the age groups, with the older benefit recipients being less likely to work (and more likely to be fully disabled). Consequently, the age difference in mean earnings is in the opposite direction in the two periods. These period differences in the gaps in the *levels* of employment and earnings between the age groups do not invalidate the DADID identification strategy. We examine whether there is any sign of the age-specific *trends* diverging up to the implementation of the reform in the next sub-section.

For both age groups, mean incomes from social transfer programs other than DI are higher at the start of the reform period than at the start of the non-reform period, and the age gap is somewhat wider in the reform period. The increase over time may well be due to the rise in the proportion of benefit recipients with mental health problems, who tend to be more heavily dependent on welfare. Combined with recipients with musculoskeletal conditions, they are the majority in all age groups and periods, and more so in the later period. In the earlier period, there is no age difference in the fraction of recipients with either of these two more subjectively defined conditions. But in the later reform period, recipients in the younger group are more likely to have these diagnoses. This gives further reason to perform disaggregated analysis by diagnosis.

4.3 Trends

Figure 1 shows difference-in-differences in receipt of any DI benefits, employment and labor earnings between the two age groups within each period.[15] These figures are drawn using monthly data to allow more detailed assessment of the evolution of the trends before and after the start of the reassessments. Each line traces the age group difference (30-44 years

[15]See Appendix B Figure B1 for plots of the raw trends in the outcomes for the two age-groups separately in the two periods.

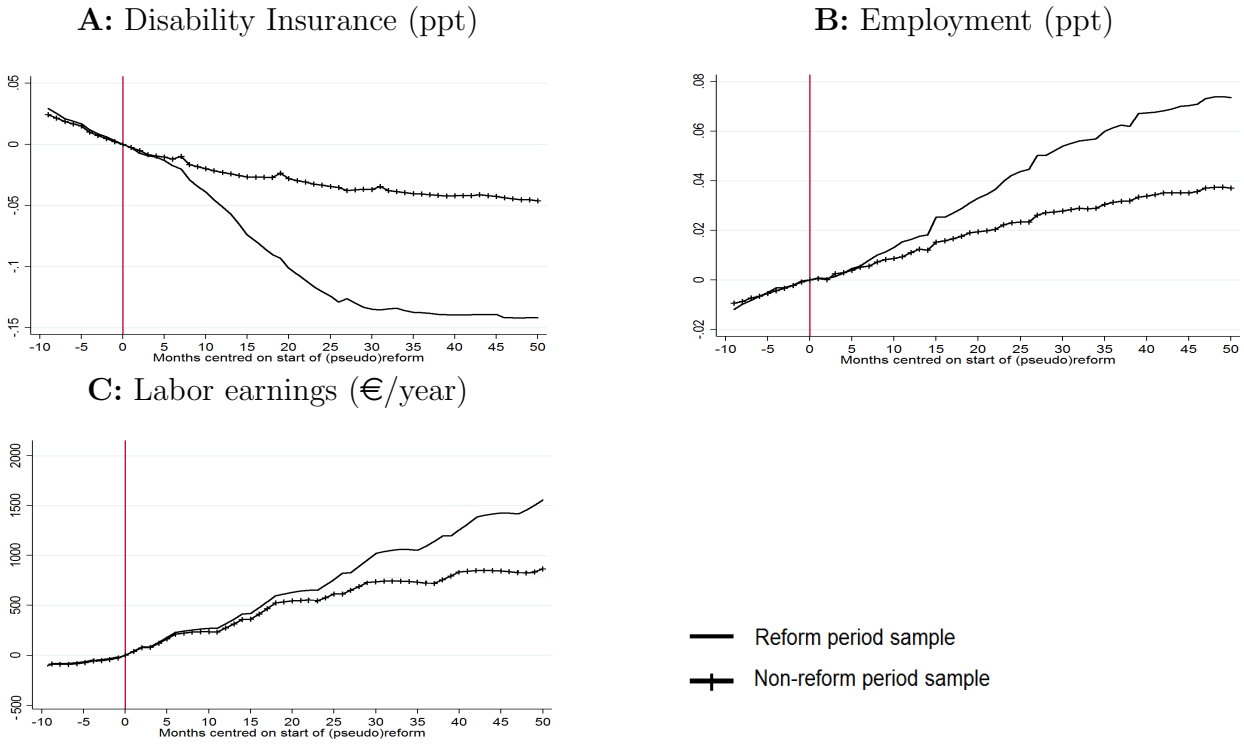
Table 3: Characteristics of DI recipients by period and age - Means at sample entry

	Reform period		**Non-reform period**	
	Age 30-44	**Age 50-53**	**Age 30-44**	**Age 50-53**
Demographics				
Female	60.3%	45.7%	53.4%	37.4%
Age	38.7	52.1	38.8	52.1
Disability insurance				
Benefit amount (€/year)	8,422	9,950	8,559	10,634
Fully disabled	63.5%	64.0%	65.4%	69.4%
Claim duration (years)	5.44	9.52	5.90	9.96
Labor market				
Employed	35.9%	35.8%	40.7%	34.6%
Earnings (€/year)	4,207	5,162	4,947	4,879
Other social transfers				
Benefit amount (€/year)	1,043	726	724	555
Diagnosis				
Mental disorders	43.1%	33.8%	34.4%	27.9%
Musculoskeletal	28.9%	32.9%	25.0%	31.2%
Other disabilities	28.0%	33.3%	40.6%	40.9%
Number of Observations	160,194	94,404	139,524	102,464

Note: The Reform period panel refers to DI benefit recipients selected in January 2004. The Non-reform period panel refers to those selected in January 1999. Columns within each panel are split by age on July 1, 2004 (Reform period) and July 1, 1999 (Non-reform period). The first column in the Reform period panel corresponds to the treatment group. All others are for comparison groups. Earnings and benefit amounts are annual, net of taxes and inflated to 2015 price levels (Eurostat Netherlands HCPI 2015).

- 50-53 years) in the deviation of the respective outcome from its value in month 0, which is October 2004 in the reform period, when reassessments started, and October 1999 in the non-reform period. After month 0, the difference in the DID between the periods corresponds to the DADID and gives an initial impression of the impact of the reform.

Figure 1: Age group difference-in-differences in outcomes by period



Note: Reform period (Jan. 2004-Dec. 2008) sample consists of individuals aged 30-44 & 50-53 on July 1, 2004 who were claiming DI in January 2004. Non-reform period (Jan. 1999-Dec. 2003) sample consists of individuals aged 30-44 &50-53 on July 1, 1999 who were claiming DI in January 1999. Month 0 is October 2004 for reform period and October 1999 for non-reform period. Each line traces a period-specific difference-in-differences: the mean outcome at month t minus the mean outcome at month 0 for the 30-44 age group less the respective difference for the 50-53 age group. Disability Insurance is an indicator of receipt of any DI benefits. Group sizes are given in Table 3. ppt = percentage points.

Consistent with the identification assumption, prior to month 0 the age group difference in the trend of each outcome is very similar across the two periods. In fact, up to month 5, i.e. five months after reassessments started in the reform period when only 8% of claimants aged 30-44 had been reassessed, there is little sign of the age differential in the trends differing across the periods. After that point, when the pace of reassessments picked up in the reform

period, the age differentials begin to diverge more markedly across the periods. This is consistent with the application of more stringent eligibility criteria to ever greater numbers of younger benefit recipients in the reform period having raised the rate at which they exited DI relative to older recipients, and with relative increases in the employment and earnings of younger recipients who either left DI or remained on the program despite experiencing a cut in their benefits.

Attribution of the differential trends across periods that are evident in Figure 1 to the reform rests on assumption (1) - the age differential in the outcome trend would have been common across periods in the absence of the reform. It is difficult to gauge the plausibility of this assumption from comparison of the outcome trends over two periods of only nine months (Jan.-Sept. 1999 and Jan.-Sept. 2004). To better assess whether the assumption is credible, we examine two different cohorts of DI claimants over a longer duration prior to the start of reassessments in the reform period. One of these cohorts consists of individuals who were: a) receiving DI in January 2003, b) aged 30-44 or 50-53 on July 1, 2004, and c) observable until December 2006. Those in the younger group of this cohort were subject to reassessment from October 2004, provided they were still on DI at that time. They are observed for 21 months prior to this date. The second cohort is defined exactly as the non-reform period groups we use for estimation except that the age criteria are applied on July 1, 2000 (rather than July 1, 1999) and we follow them only until December 2002. The pseudo reform period for this cohort is set as starting in October 2000.

Figure 2 shows the age group differential in the trends in the receipt of any DI benefits, employment and labor earnings over the four years that these cohorts are followed. Over the 21 months prior to the start of the reassessments of the reform period sample, the age differentials in the outcome trends do not diverge markedly between the two cohorts. This is slightly less true for the receipt of DI benefits than it is for the other two outcomes. Apparently even before the start of reassessments in the reform period sample, younger claimants in this cohort were exiting DI at a faster rate relative to older claimants than was

Figure 2: Age group difference-in-differences in outcomes by period - extended duration prior to (pseudo) reform



Note: Reform period (Jan. 2003-Dec. 2006) sample consists of individuals aged 30-44 & 50-53 on July 1, 2004 who were claiming DI in January 2003. Non-reform period (Jan. 1999-Dec. 2002) sample consists of individuals aged 30-44 & 50-53 on July 1, 1999 who were claiming DI in January 2000. Month 0 is October 2004 for reform period and October 2000 for non-reform period. Sample sizes are 140,283 for the non-reform period sample claimants aged 30-44, and 103,490 for those in the same period aged 50-53. In the reform period, the sample size of the treatment group is 155,973, and it is 92,298 for claimants aged 50-53.

the case in the earlier period sample. This would be consistent with recipients in the later period leaving the program in anticipation of negative reassessments. This seems unlikely given there is no sign of a similar pre-reform divergence in the employment trends. Someone who anticipated that their DI benefits would be terminated or cut would have no incentive to leave the program before this occurred, unless they had found employment. There is a clear downward kink in the differential trend in receipt of DI in the reform period sample coincident with the acceleration in the reassessments from around month 5 and no such kink in the non-reform period sample. The size of this divergence relative to the prior differential trend suggests that while the DADID may overestimate the impact of the reform on the DI exit rate, the upward bias is likely to be small. Further, the similarity of the trends in employment and earnings prior to month 0 across periods supports the validity of the DADID identification assumption for these outcomes.

5 Results

5.1 Main estimates

Column (1) of Table 4 gives the estimate of β_4 from a least squares regression of the form (3) for each outcome. Each column entry is a DADID estimate of the ATET - the effect of the reform on the respective outcome in 2008 averaged over all individuals who were aged 30-44 and claiming DI in 2004. By 2008, these individuals had been subjected to reassessment under the more stringent criteria. The middle column gives a measure of the relative impact of the reform. It shows each estimated effect scaled by the treatment group's predicted mean outcome in 2008 under the counterfactual of no reform.[16] Column (3) gives effects on labor market outcomes and other social transfer income scaled by the estimated effect on

[16]The counterfactual is obtained by subtracting the estimated effect from the mean of the treatment group's (regression) predicted outcome in 2008, i.e. $\frac{1}{n_T}\sum_i 1\left(AGE_i \times PERIOD_i \times YEAR_4\right)\hat{Y}_{it} - \hat{\beta}_4$, where $\hat{Y}_{it}$ is the predicted outcome from (3), 1() is the indicator function and n_T is the number of individuals in the treatment group. The estimates are given in Appendix C Table C1.

DI income, which facilitates comparison of the sizes of the responses induced by the 2004 Dutch reform with those generated by other policies that lead to changes in DI benefits.[17]

We estimate that this reduced the probability of remaining on DI in 2008 by 14.4 percentage points. This includes both the direct effect of claims terminated through application of the stricter rules and any indirect effect that may arise through reduced benefits inducing some to leave DI. Even without the cuts, some claimants would have left the program by 2008. Using the regression estimates, we predict that 84.5% of individuals aged 30-44 who had been claiming DI in 2004 would still have been on the DI roll in 2008 if there had been no tightening of the rules. This implies that reassessment with stricter criteria reduced the probability of continued receipt of DI by 17% of what it otherwise would have been. It raised the DI exit rate by 93%. On average, reassessment is estimated to have reduced the annual amount of DI benefit received by €1565, or around one fifth of the average amount under the counterfactual.[18] Given that the degree of disability did not change as a result of reassessment for a majority and it even increased for a few (Table 2), this average grossly understates the average reduction in benefits experienced by the 34% for whom the outcome of reassessment was negative. To estimate this reduction, we need to make an assumption about its magnitude relative to the size of the effect on the small proportion who had their degree of disability raised following medical reexamination (due to health deterioration) despite application of more stringent rules.[19] If the magnitudes of the two effects were equal,

[17] We refer to these as "scaled effects", rather than instrumental variables (IV) estimates of the response of labor outcomes to DI benefits, for three reasons. First, it is possible that reassessment could impact on labor activity other than through benefit entitlement, and so the exclusion restriction could be violated. Second, the estimated reduction in benefits is the combined effect of cuts and responses to those cuts through claimants leaving DI because it has become less generous. Third, reassessment resulted in benefit entitlement rising for some recipients whose heath had deteriorated sufficiently to offset the effect of increased stringency. Hence, monotonicity does not hold.

[18] We estimate that reform reduced the rate at which DI income replaced pre-disability earnings by 7.2 percentage points from a replacement rate under the counterfactual of 46 percent. To obtain these estimates, we average the replacement rate over the whole treatment group and set it to zero for those who had left DI by 2008.

[19] We can write the ATET as a weighted average of the effects on the sub-groups that have their benefits cut and raised: $ATET = p_c ATET_c + p_r ATET_r$, where $ATET_c = \mathbb{E}\left[Y_{i4}^1 - Y_{i4}^0 \mid D_i = 1, Y_{i4}^1 < Y_{i4}^0\right]$, $ATET_r = \mathbb{E}\left[Y_{i4}^1 - Y_{i4}^0 \mid D_i = 1, Y_{i4}^1 > Y_{i4}^0\right]$, p_c is the proportion of the treated who have their benefits cut $\left(p_c = \frac{\sum D_i 1(Y_{i4}^1 < Y_{i4}^0)}{\sum D_i}\right)$ and p_r is the proportion for whom benefits are raised. Let $-ATET_r = kATET_c$, then

then the average benefit reduction of €1565 over all those reassessed would imply an average reduction of €5530 among those whose benefits were cut. This is probably an overestimate. But even if we assume that there was no effect on the 6% whose degrees of disability were raised, then the average effect on the 34% whose benefits were cut would still be a substantial €4549.[20] This is 54% of the mean benefit income received by the treatment group prior to the reform.

These estimates confirm that the 2004 reform substantially reduced DI benefits and dependency. It was about twice as aggressive as the reassessment of Dutch DI claimants a decade earlier that is estimated to have lowered the probability of remaining on DI by 3.8 percentage points and reduced the amount of benefit received by 10% (for claimants aged 45) (Borghans et al. 2014).

Having established that the reform substantially reduced DI benefits, we now turn to the question of central interest: what impact did this increased stringency have on employment and earnings? We estimate that reassessment raised the probability of employment by 6.7 percentage points, which is a 20% increase relative to the predicted employment rate in the absence of the reform and is equivalent to a 4.3 point rise in employment for a €1000 cut in annual benefits (Table 4). To determine whether any of this increase is due to movements into work among those remaining on DI, we estimate the effect on the probability of working and not claiming and find a significant increase of 8.5 points (SE=0.18, p-value<0.01). The larger magnitude of this effect implies that reassessment actually reduced the likelihood of claiming DI and working (by 1.8 points), which is likely due to partially disabled working claimants be forced or induced to leave the program.

$ATET_c = \frac{ATET}{p_c - kp_r}$. We assume the average treatment effect is zero for recipients whose degree of disability remained the same after reassessment.

[20]There are two reasons to expect the magnitude of any effect on recipients who had their degree of disability (DD) increased to be small, possibly zero, and, in any case, substantially smaller than the effect on those whose DD was reduced. First, any increase in benefit entitlement due to health deterioration would be (partially) offset by using more stringent rules to calculate DD. Second, target group recipients with deteriorating health, along with equivalent cases in the comparison groups, may have been detected eventually by the periodic reassessments that were conducted prior to the 2004 reform. Then, subject to our identification assumption, the empirical strategy would give a zero effect on these recipients.

Table 4: Effects of reassessment of DI recipients under more stringent rules

	Effect	Effect scaled by predicted mean if no reform	Effect scaled by benefit reduction (in €'000s/year)
	(1)	(2)	(3)
Disability Insurance			
Benefit Receipt (ppt)	-14.40***	-17.0%	NA
	(0.17)		
Benefit Amount (€/year)	-1,565***	-19.8%	NA
	(31.70)		
Labor Market			
Employment (ppt)	6.68***	19.8%	4.27
	(0.22)		
Days worked (year)	17.03***	22.3%	10.88
	(0.58)		
Earnings (€/year)	995***	18.1%	635.8
	(43.19)		
Other social transfers			
Benefit amount (€/year)	376***	42.9%	240.3
	(17.73)		
Number of individuals	496,586		
Number of observations	2,482,930		

Notes: Column (1) gives least squares estimates of β_4 from (3). Standard errors, in parentheses, are adjusted for clustering at the individual level. Column (2) gives the column (1) estimate as a percentage of the predicted mean outcome of 30-44 year olds in 2008 under counterfactual of no reform (see footnote 16). See Appendix C Table C1 for these predictions. Column (3) gives the column (1) estimate divided by the absolute value of the estimated effect on the benefit amount in €'000s (from 2nd row of column (1)). The number of individuals is the total across all treatment and comparison groups. For the numbers in each group, see Table 3. ppt = percentage points. *** indicates significance at the 1% level.

Borghans et al. (2014) estimate that the less stringent tightening of the Dutch DI program in 1993 increased employment by 2.9 points. In absolute terms, this is less than half the size of the effect we find on employment. But it is larger relative to their estimated 3.8 percentage points reduction in the probability of receiving DI. The implied lower rate of absorption of displaced claimants into employment from the later reform we evaluate is consistent with an expected decrease in the work capacity of claimants as the process of DI retrenchment proceeds. Moore (2015) estimates that 22 percent of US DI recipients whose benefits were terminated entered employment. Set against the 100 percent loss of benefit entitlement, in relative terms this is a much smaller employment effect than we find. This might be due to the addictive behavior of the group targeted by the reform evaluated by Moore, as well as incentives for disqualified US claimants to stay out of work to strengthen their case at reapplication. There is no such incentive in the Dutch system.

We estimate that greater benefit stringency increased the number of days worked annually by 17; equivalent to 22% of the predicted mean for the treatment group in the absence of the reform. The extensive and intensive margin effects on labor supply produced an estimated €995 average increase in the annual earnings of DI claimants whose entitlement was reassessed. This is an 18% increase relative to predicted earnings under the counterfactual. It is almost two thirds of the estimated average reduction in the benefits received. From a €1000 reduction in benefits, €636 could be regained through labor market earnings.[21] This is very close to the estimate of Borghans et al. that earnings rose by 61 cents for each €1 reduction in benefits resulting from the 1993 reassessment of Dutch DI claimants. Apparently, even after that reduction in entitlement and a 2002 reform that is likely to have reduced the

[21]Bearing in mind that the estimated reduction in benefits arises not only directly from cuts imposed after reassessment but also indirectly from decisions to leave DI that has become less generous, the ratio of the estimated effects on earnings and benefit income cannot be interpreted as an unbiased estimate of the rate at which earnings are crowded out by each €1 of DI benefit. However, we can infer that the rate of crowd-out is at least as high as 0.64:1, since the average imposed cut in benefits will be less than the average reduction in benefits received. Also note that we are taking the ratio of two averages, not the average of a ratio. However, given that the expected effect on earnings relative to the expected effect on benefits is a lower bound of the expectation of the ratio of the individual earnings effect to the individual benefit effect (Cochran 2007), it is anticipated that we underestimate the average degree to which earnings replace lost benefit income.

rate of entry to DI (Koning and Lindeboom 2015, see section 5.3), DI recipients subjected to reassessment in 2004 still had considerable earnings capacity they could draw on to replace a substantial part of the benefits lost due to the increased program stringency. This is even more striking considering that those affected had been claiming DI for more than five years, on average, and 63% were classified as fully disabled (see Table 3).

It bears emphasis that these are average effects and reassessment resulted in the termination of benefits, or their reduction, for a little more than one third of recipients (Table 2). If we assume that reassessment did not have any impact on earnings other than through benefit entitlement and it had no effect on the earnings of the 6% whose degree of disability was raised, then an average increase in earnings of €995 over all those reassessed implies an average increase of €2892 over all those who had their benefits cut as a result of reassessment.[22] This is 69% of the average annual earnings of the whole treatment group prior to the reform and is a 53% increase on the predicted mean earnings in 2008 if there had been no reform. These large average effects do not, however, reflect the predicament of claimants negatively impacted by reassessment who could not increase their earnings to an extent anywhere near sufficient to achieve the average 64% replacement of lost benefit income.

We estimate that reducing DI entitlement increased the amount received from other social transfers by €376, on average (Table 4).[23] This is 24% of the average reduction in income received from DI. The analogous estimate from Borghans et al. (2014) is 30%. Apparently, opportunities to substitute between programs decreased in the decade between the reforms evaluated, but not markedly. Summing the average effects on earnings and other

[22] In addition to the reasons given in footnote 20 for expecting the magnitude of any effect on the benefit entitlement of recipients whose degree of disability (DD) was increase to be small, and possibly zero, the effect on their earnings would be even smaller relative to that on those whose DD was reduced if, as seems likely, the earnings response to a benefit increase (due to worsening health) is smaller than that to a benefit reduction (with constant health). Using the formula given in footnote 19, if we assume the earnings effect on those whose DD was raised is one tenth of the size of the effect on those whose DD was reduced, then the average earnings effect on the latter group would be €2944. If we assume equal but opposite effects on the two groups, then the effect on those whose benefits were cut would be €3516. In any case, the effect on those who experienced a cut in benefits appears to have been substantial.

[23] Around half of the spillover to other programs was to unemployment insurance (UI) (Appendix C Table C2). Those deemed ineligible for DI were automatically transferred to UI if they had made sufficient social insurance contributions prior to entering DI.

social transfer income gives a total of €1371, which is about 88% of the estimated average reduction in payments received from DI. Without taking this compensation into account, on average, the cuts to DI benefits lowered income from all sources by 11% relative to what it would have been if there had been no reform.[24] The relative drop in income becomes 1.3% when compensation through earnings and other social programs is taken into account.

5.2 Placebo test

The validity of our empirical strategy rests on the assumption that the age differential in the outcome trends that would have materialized between 2004 and 2008 in the absence of the DI reform is that which occurred between 1999 and 2003. To further assess the plausibility of this assumption, we perform a *placebo test* by estimating the DADID in outcomes of individuals who were not recipients of DI benefits, and so were not exposed to the reform, but who were potentially affected, possibly differentially by age, by differences in labor market conditions across the two periods. Placebo treatment and comparison groups are defined by age and period analogous to those used to estimate the effect of the reform. The difference is that we only use individuals who did not claim DI at any time between January 2004 and December 2008, and in the non-reform period between January 1999 and December 2003. The placebo treated individuals are the same age as the DI recipients who were reassessed (30-44 on July 1, 2004) and are observed between 2004 and 2008. The placebo comparison groups are (1) 50-53 on July 1, 2004, (2) 30-44 on July 1, 1999 and (3) 50-53 on July 1, 1999. We exclude individuals who were claiming unemployment insurance in 1999 (for non-reform period groups) or 2004 (for reform period groups) because the DI reform could potentially have affected their labor market opportunities by increasing the supply of labor from DI claimants. After imposing these exclusion restrictions, there are 6.7 million individuals available for the analysis. We use a random 50% sample of them.

[24]This percentage is calculated using the average reduction in DI income from column (1) of Table 4 and the predicted counterfactual means of DI income, earnings and other social transfer income given in Appendix C Table C1.

Table 5: Placebo test - empirical strategy applied to data on non-recipients of DI

	Effect (1)	Predicted mean (2)	Effect scaled by predicted mean (1)/(2) × 100
Employment (ppt)	-0.57***	73.43	-0.78%
	(0.01)		
Days worked (year)	-0.15	217.18	-0.07%
	(0.15)		
Earnings (€/year)	-195.90	34,061	-0.58%
	(125.78)		
Number of individuals	3,345,789		
Number of observations	16,728,945		

Notes: Column (1) gives least squares estimates of β_4 from (3) using individuals who did not claim DI at any time in the respective period (1999-2003 or 2004-2008) and were not claiming UI at the beginning of the period. Column (2) gives predicted mean outcome of 30-44 year olds in 2008 assuming the age differential in the outcome trend 2004-2008 equals that observed 1999-2003 (see footnote 16). Right-hand column gives the estimate in column (1) as a percentage of the prediction in column (2). Standard errors, in parentheses, are adjusted for clustering at the individual level. The number of individuals is the total across all placebo treatment and comparison groups. ppt = percentage points. *** indicates significance at the 1% level.

The results presented in Table 5 show precisely estimated zero "effects" on two of the three labor market outcomes. There is a very small, but statistically significant, negative "effect" on employment.[25] Given the size of the estimate, its significance may simply be attributable to the huge sample. The estimate suggests that employment of individuals aged 30-44 who were not recipients of DI *fell* by only 0.8% of what it would have been in 2008 if the age differential in the employment trends between 2004 and 2008 had been the same as that observed between 1999 and 2003. Under the same assumption, we estimate that the DI reform *raised* employment of DI recipients aged 30-44 by 20%. Hence, if anything, we may be slightly underestimating the impact on employment. But the placebo test suggests that any such bias is marginal, and it gives no reason to doubt the validity of the identification with respect to the effects on earnings and days worked.

5.3 Robustness

The placebo test indicates little or no bias arising from differences in labor market conditions across the two periods that may have affected age groups differently. A second potential threat to the identification would be any change in DI prior to the 2004 reform that had a different impact on older and younger benefit recipients. One change in DI that occurred within the estimation periods was the introduction of the so-called Gatekeeper Protocol (GP) in 2002. This made the employer and the employee jointly responsible for taking active measures to enable the latter to continue working during the waiting period for DI. It is credited with substantial reductions in the rate of DI inflow (De Jong et al. 2011; Van Sonsbeek and Gradus 2012; Koning and Lindeboom 2015). Any impact on the DI exit

[25]The direction of this effect may seem puzzling given that macroeconomic conditions were better in 2004-2008 than they were in 1999-2003. But this is not simply a period effect. It is an age difference in the period effect on the trend. To make this explicit, we decompose the estimate by running (1) a DID regression across the two periods using individuals aged 30-44 in each, and (2) a DID regression across the two age groups using individuals observed in the 2004-2008 period. The latter produces an estimated negative effect on the employment of 30-44 year-olds: their employment improved by less than that of 50-53 year-olds between 2004 and 2008. DID (1) produces an estimated positive effect on employment: the employment of 30-44 year-olds increased by more in the 2004-2008 period than it did in the 1999-2003 period. This is consistent with the positive turn of the business cycle from 2004.

rate, as well as on the employment and earnings of those already receiving DI, which are relevant here, would be indirect. And there is no reason to expect that any such effect would differ between older and younger benefit recipients. However, if there were age heterogeneity in the impact on the entry rate, then this could possibly bring about a difference across our reform and non-reform periods in the relative composition of the age groups of DI recipients. If this happened, it could potentially induce bias.

The GP affected claimants who entered DI in January 2003 and later. It is irrelevant to our non-reform period sample, who are selected from the stock of DI recipients in January 1999, and to all in the reform period sample except those with a claim duration of 12 months or less in January 2004, when we select this sample from the stock of DI recipients. We test whether the GP is confounding our estimates by dropping all DI recipients who had been claiming for 12 months or less at the time of selection into either the reform period or the non-reform period panels. The estimated effects on DI benefits and employment given in panel B of Table 6 are very close to the respective estimates obtained from our main design, which are reproduced in panel A. The effect on the probability of receiving DI is about two percentage points smaller than the main estimate and the effect on earnings is about one fifth smaller. With this restriction on the samples, we estimate that reassessment that reduced benefit entitlement by €1000 would raise earnings by €534, compared with a main estimate of €636. These differences could indicate some upward bias in the earnings effect of the 2004 reform arising from changes in the composition of the stock of DI recipients brought about by the GP. But they could also reflect heterogeneity in the response to the reform by claim duration, which we explore in section 5.4. In any case, the main conclusion from this analysis is that it does not appear that the GP, rather than the 2004 reform, is driving our results.

Our choice of the 50-53 age range to define the older comparison group is motivated by a compromise between keeping reasonably close to the age of the treatment group and obtaining a large sample (for heterogeneity analysis). Panel C of Table 6 provides estimates

Table 6: Robustness to alternative sample selections and use of monthly data

	Disability Insurance		**Labor Market**			
	Benefit Receipt	**Benefit Amount**	**Employment (ppt)**		**Earnings (€/year)**	
	(ppt)	(€/year)	Effect	Scaled effect	Effect	Scaled effect
	(1)	(2)	(3)	(3)/|(2)| × 1000	(5)	(5)/|(2)| × 1000
A. Main estimates						
	-14.40***	-1,565***	6.68***	4.27	995***	636
	(0.17)	(31.7)	(0.22)		(43.2)	
B. Drop those with claim duration ≤ 12 months						
	-12.50***	-1,504***	6.85***	4.55	803***	534
	(0.20)	(33.5)	(0.25)		(53.7)	
C. Define comparison group by other ages						
Ages 50 to 52	-14.20***	-1,615***	6.90***	4.27	968***	599
	(0.21)	(39.7)	(0.27)		(58.1)	
Ages 50 to 54	-14.10***	-1,584***	7.03***	4.44	990***	625
	(0.19)	(33.4)	(0.24)		(49.8)	
D. Use monthly data						
	-13.57***	-1,521***	5.67***	3.73	784***	515
	(0.37)	(65.2)	(0.46)		(93.1)	

Notes: Panel A reproduces the main estimates from Table 4 obtained using annual data on the stock of recipients in January 2004 (reform period) and January 1999 (non-reform period) with the older comparison group defined by the age interval 50-53. Panel B removes recipients with a claim duration of 12 months or less at entry to the panels. Panel C redefines the older comparison group by the age intervals 50-52 (top row) and 50-54 (bottom row). Panel D estimates are obtained using monthly data. In this case, differences are taken relative to September 2004 (in reform period) and estimated effects at December 2008 are presented. Sample sizes (number of individuals): Panels A & D = 496,586, Panel B = 447,5443, Panel C (top row)= 443,196, Panel B (bottom row)=525,957. To get number of observations, multiply number of individuals by 5 for Panels A-C and by 60 for Panel D. For other details see Notes to Table 4.

using narrower and wider age intervals to select the comparison group. They are very similar to the main estimates.

As acknowledged in section 3, using annual data and taking differences from 2004 introduces a slight inaccuracy because 1% of reassessments were carried in the last quarter of that year. Given this fraction is very small and, in any case, there was a lag of a few months between reassessment and benefit cuts taking effect, this is unlikely to cause any bias that is not negligible. However, while effectively all recipients aged 30-44 had been reassessed by the end of 2008, around 3% were reassessed during that year (Table 1). The full effect of reassessment on these recipients may not be reflected in earnings averaged over 2008. To allow for both inaccuracies, we test robustness to using monthly data that allow us to take differences between September 2004 and December 2008.[26] The estimates obtained (Table 6, Panel D) are reasonably close to the main estimates. The effects on receipt of DI and employment are both about one percentage point smaller than the respective main estimates, and the estimated effect on earnings is €210 smaller. But, as with the results of the other robustness analyses, these differences do not change the general conclusions that reassessment resulted in substantial cuts to DI benefits, and both employment and earnings increased markedly to compensate.

5.4 Effects by duration of claim

Having established that Dutch DI recipients had considerable earnings capacity that could be tapped to replace a large fraction of the benefit income lost as a result of reassessment under more stringent rules, we now turn to the question of whether earnings capacity diminishes while claiming DI. If so, then evidence indicating that rejected DI applicants have substantial earnings capacity (Autor et al. 2017; Chen and Van der Klaauw 2008; French and Song 2014; Maestas et al. 2013; Von Wachter et al. 2011) is a poor guide to the effectiveness of measures

[26] We cannot estimate effects after December 2008 since this would require extending the length of the non-reform period, which cannot start before January 1999 due to data not being available. If the non-reform period where extended in the other direction, then the younger comparison group would then become exposed to the reform at the end of this period.

that aim to raise the exit rate from a DI program without causing undue hardship by withdrawing the benefits of claimants with little scope to substitute with earnings.

Following Moore (2015), we extend the regression (3) to allow the effect of the reform on earnings to vary with the duration of a claim. Let $DURATION_i$ be the number of years an individual had been receiving DI benefits at entry to the sample (January 1999 for the non-reform period and January 2004 for the reform period). We interact a third order polynomial of this duration with the product of the age group, period and year indicators that identifies exposure to the reform and so allow the effect of reassessment to vary nonlinearly with the duration of a claim.

$$Y_{it} = \sum_{t=1}^{4}\sum_{j=0}^{3}\Big[\beta_{tj}\left(AGE_i \times PERIOD_i \times YEAR_t \times DURATION_i^j\right) + \theta_{it}\left(YEAR_t \times DURATION_i^j\right) + \gamma_{it}\left(AGE_i \times YEAR_t \times DURATION_i^j\right) + \delta_{it}\left(PERIOD_i \times YEAR_t \times DURATION_i^j\right)\Big] + \mu_i + \varepsilon_{it} \tag{4}$$

If the DADID identification assumption (1) holds conditional on each duration of a claim, then the average treatment effect on those exposed to the reform and its nonlinear variation with duration are obtained from the β_{4j} parameters. β_{40} corresponds to the effect of reassessment on someone who had just entered DI in January 2004. Adding combinations of the parameters β_{41}, β_{42} and β_{43} gives effects at positive durations.

People who claim DI for different lengths of time also differ in other respects. In both the reform and non-reform periods, those who had been claiming DI for longer at entry to the sample are older and are more likely to be female (see Appendix B Figure B2). The faction of fully disabled benefit recipients rises with claim duration up to about 10 years, then falls (Figure B2). We document heterogeneity in the effect of the reform by these characteristics in sections 5.5 and 5.4.1. This heterogeneity potentially confounds that by claim duration. Hence, we also estimate (4) separately by demographic and disability groups, which has the added advantage of revealing any heterogeneity that exists across these groups in the way

in which the effect varies with claim duration.

Using the full sample without disaggregation, estimates of parameters β_{41}, β_{42} and β_{43} for each of the outcomes are separately and jointly statistically significant (Appendix C, Table C3), implying that the effects do indeed vary with the duration of a claim. This variation is traced in Figure 3. We show estimates up to a duration of 15 years, which corresponds to the 80^{th} percentile in the treatment group. The effect on the probability of receiving DI benefits and on the amount received both initially rise in magnitude as the length of the claim prior to the reform increases up to about 6 six years. Thereafter, both effects fall in magnitude. Note that we are estimating the effect of reassessment of degree of disability, which is the ratio of assessed earnings capacity to previous earnings and determines benefit entitlement. The larger effect on the amount of DI benefits paid to those who had been claiming up to 6 years (Panel B) implies that medical reexamination and application of more stringent entitlement rules resulted in greater upward revision to the earnings capacity as duration lengthened up to this point. From then on, the revision to earnings capacity was smaller for those who had been claiming for longer, which is consistent with deterioration in earnings capacity during the time someone has been claiming DI offsetting the effect of imposing stricter criteria.

The effect on employment (Panel C) changes little up to a claim duration of about two years, and then declines rapidly with the length of a claim, becoming insignificantly different from zero just above 8 years and bottoming out at around 12 years. This pattern does not resemble the inverted U-shaped employment response to DI terminations that Moore (2015) finds with US data. In that study, the response increases by about 8 percentage points up to a claim duration of three years. Moore suggests this could be due to a rehabilitation effect: time on DI provides an opportunity to recover health and work capacity. We find little or no evidence consistent with this hypothesis. Besides the obvious institutional differences, the discrepancy may be partly due to the different nature of the reforms examined. Moore estimates the response to a homogeneous treatment: the termination of DI benefits. In our case, while all claimants were subject to the same reassessment process, the consequences

for their benefits differed, as is clear from the U-shaped relationship in Panel B.

The effect on earnings also declines steadily with claim duration (Panel D). There is less evidence of the effect bottoming out in this case and it remains significant even after 15 years. The falling effects on employment and earnings occur despite the increasing reduction in benefits up to a claim duration of around 6 years. Figure 4) shows the employment and earnings effects scaled by the effect on the benefit amount. Both scaled effects decline continuously with claim duration. Those who had been claiming for less than a year when benefits were cut were able to overcompensate for lost benefit income with labor earnings. But those who had been claiming for 6 years or more were able to recover less than half of more modest cuts in benefits with increased earnings.

5.4.1 Effects by cause and degree of disability

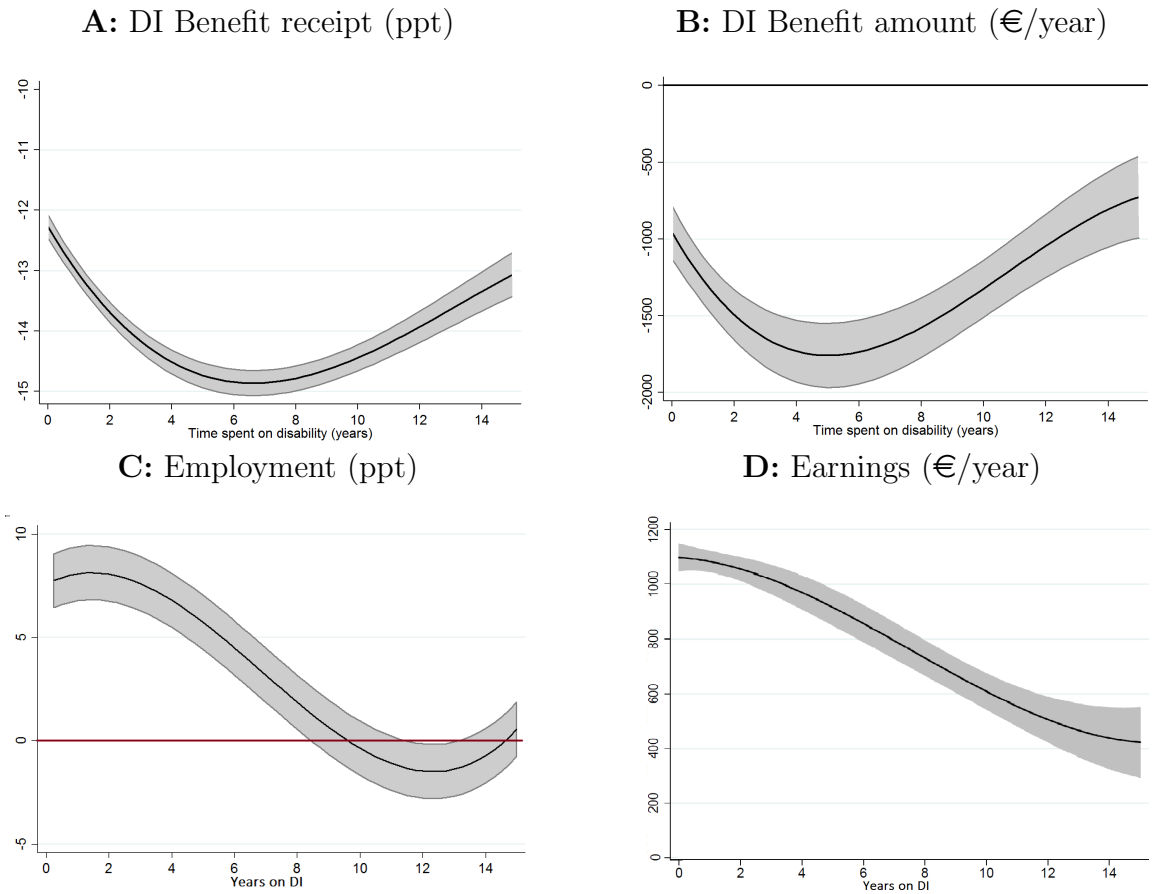
To assess the hypothesis that benefit recipients who qualified through more subjectively defined and difficult to verify conditions have the greatest earnings capacity, we split the sample into three groups according to main diagnosis at entry to DI—musculoskeletal disorders, mental health disorders and a residual category of all other disorders[27]—and estimate the regression model (3) separately for each sub-sample.[28] The estimates given in the top panel of Table 7 show that, consistent with the hypothesis, the employment response to reassessment is largest for those with musculoskeletal disorders, followed by those with mental health disorders.[29] This is entirely due to recipients with these conditions being much more likely to exit DI after reassessment. Their absorption into work appears to have been substantially lower. The employment effect is around 40% of the impact on the DI exit rate for claimants with musculoskeletal and mental health conditions. For those in the residual

[27]The classification uses the most aggregated level of the International Classification of Diseases version 9. The residual category includes diseases and conditions related to the nervous system, many of which, such as physical pain, are difficult to verify. Disaggregated analysis revealed that the response to the reform in this category did not resemble the responses in the mental health and musculoskeletal categories more than it resembled the response of recipients with the remaining diseases included in the residual category.

[28]All results given Tables 7 and 8 are obtained by splitting the sample and estimating regression (3) separately for each group.

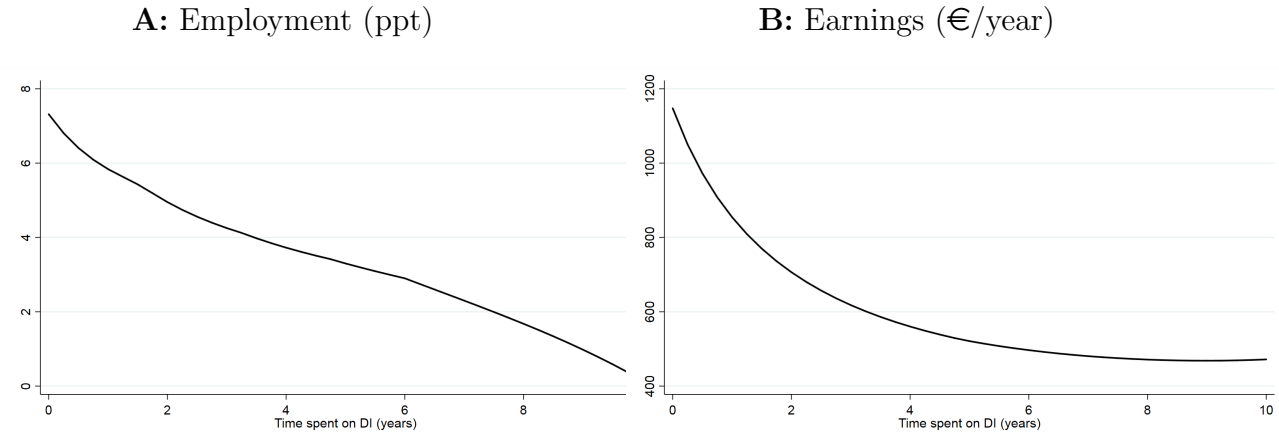
[29]These differences, like all other heterogeneous effects referred to in the text, are significant.

Figure 3: Effects of reassessment of DI recipients under more stringent rules by duration of claim



Notes: Estimates of average effects of reassessment by length of time on DI at entry to the sample. Derived from least squares estimates of $\beta_{4j}, j = 0, 1, 2, 3$ from (4) ((Appendix C Table C3). Shading indicates 95% confidence interval computed from delta method standard errors adjusted for clustering at the individual level. ppt = percentage points. Sample size is 496,586 individuals and 2,482,930 observations.

Figure 4: Employment and earnings effects scaled by effect on benefit amount (in €1,000) by duration of claim



Notes: Panel A shows employment effects reported in Panel C of Table 3 scaled by the estimated effects on benefit amount reported in Panel B of Table 3 (in €1000). Panel B shows earnings effects reported in Panel D of Table 3 scaled by the estimates reported in Panel B of Table 3. See Notes to Table 3.

category of presumed more objective, verifiable disorders, the increase in employment is 70% of the fall in DI participation. This discrepancy could be taken as indicative of overly strict reassessment of recipients with musculoskeletal and mental health problems that produced upwardly biased evaluation of their earnings capacity. Their benefits were terminated and yet they were not able to find work. An alternative interpretation is that the low absorption rate reflects greater distaste for work among those who had qualified for DI through these two types of conditions. This is what one would expect if the work-shy were more likely to take advantage of the difficulty of verifying the validity of a claim of incapacity to work because of a mental health or musculoskeletal problem. Unfortunately, we cannot assess the relative veracity of these two explanations, which obviously have very different implications for the welfare impact of both the reform and the DI program.

In absolute terms, the average earnings response is about twice as large for those with musculoskeletal and mental health problems as it is for those with any other type of disorder. The earnings effect relative to the predicted mean under the counterfactual (given in square brackets) is largest for claimants with mental disorders (28%) and lowest for those in the residual category (10%). This ranking is maintained when the earnings increase is expressed

per €1000 reduction in DI income (mental health - €746; all other disorders - €558). Since a larger proportion of those in the residual category who left DI entered employment, this suggests that claimants in this group were taking lower paid jobs, or they were not increasing labor supply at the intensive margin as much to compensate for the loss of partial disability benefits.[30] Overall, the results indicate that benefit recipients with mental health and musculoskeletal problems did indeed have greater scope to increase earnings in response to cuts in DI benefits, which were more aggressive for these claimants as a result of being reassessed to have greater earnings capacity.[31] Further, the earnings effect declines steeply with claim duration among recipients with the two more subjectively-defined conditions, while the effect is almost constant for those with other conditions (Figure 5, Panel A).[32] This is consistent with capacity for work and/or preferences for it deteriorating during the time spent on DI only among those who qualify through more difficult to verify health problems.

Reassessment reduced DI receipt of the initially partially disabled by twice as much as the probability of remaining dependent on DI was reduced among the initially fully disabled (Table 7, middle panel). Despite this, the employment effect on the former group is only half as large as that on the latter group. The differential impact on DI receipt is because a downward revision in the degree of disability could bring a partially disabled claimant below the 15% threshold necessary to qualify for any DI, while the same reduction might push a fully disabled claimant into partial disability. The differential impact on employment is because 68 percent of partially disabled recipients were working initially, compared with 18 percent of the fully disabled (Appendix B Table B5). Obviously, there is greater scope for those in the latter group to enter work. This is confirmed in the bottom panel of the table, which shows that reassessment raised the employment of the partially disabled who were

[30]For all heterogeneous effects on days worked and on income from other social transfers, see Appendix C Table C4.

[31]These differences do not simply reflect sex, age and degree of disability differences by cause of disability (Appendix B Table B4). The earnings responses are largest for those with mental and musculoskeletal conditions within all sex, age and degree of disability groups (Appendix C Tables C7 and C6).

[32]The steeper decline for those with musculoskeletal considitons than for those with mental health problems is partly driven by differences by sex. Recipients with musculoskeletal conditions are more likely to be male and the earnings decline with claim duration is more rapid for men (Figure 5, Panel D).

not initially working by almost as much as it raised the employment of the fully disabled. Benefit recipients who were initially fully disabled increased their earnings by more than the partially disabled both absolutely and relative to the counterfactual mean. But this is because the benefits of the fully disabled were cut by more. For a €1000 reduction in benefit income, the increase in earnings is similar (€626-674). Even those who had been classified as fully disabled were able to increase earnings to an extent sufficient to replace more than three fifths of the benefit income lost.[33]

The earnings effect of reassessment declines more steeply with time on DI among fully disabled recipients than it does among the partially disabled (Figure 5, Panel B). The fact that the effect decreases with claim duration among the fully disabled suggests that the presence of this relationship in the full sample is not simply due to recipients with longer durations are more disabled. Splitting the partially disabled by initial employment status reveals that the relationship mainly differs with the latter characteristic (panel C). The earnings effect on the partially disabled who were not initially employed falls steeply with time claiming DI. After a duration of 3.5 years, there is no significant earnings effect in this group. In contrast, reassessment induces the partially disabled who were initially working to raise their earnings, the effect is always significant and it is as strong for those who had been on DI for 15 years as it is for those who had just entered. This is consistent with partial disability combined with employment being effective in maintaining earnings capacity as the time claiming DI lengthens.

[33]The remainder of the loss of DI income was replaced by increased payments from other social transfer programs (see Appendix C Table C4).

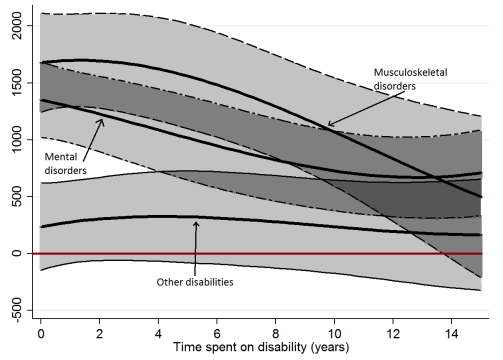
Table 7: Effects of reassessment of DI recipients under more stringent rules by cause and degree of disability

	Disability Insurance		**Labor Market**				**No. individuals**
	Benefit Receipt (ppt)	Benefit Amount (€/year)	Employment (ppt)		Earnings (€/year)		
			Effect	Scaled effect	Effect	Scaled effect	
	(1)	(2)	(3)	(3)/|(2)| × 1000	(5)	(5)/|(2)| × 1000	(7)
Cause of disability							
Musculoskeletal	-19.81***	-2,015***	7.82***	3.88	1,221***	606	144,172
	(0.32)	(58.06)	(0.42)		(83.47)		
	[23.03%]	[27.83%]	[18.84%]		[16.93%]		
Mental	-16.14***	-1,549***	6.45***	4.16	1,156***	746	177,596
	(0.27)	(51.82)	(0.37)		(66.17)		
	[18.34%]	[18.49%]	[22.22%]		[27.45%]		
Other	-7.80***	-1,111***	5.48***	4.93	620***	558	174,816
	(0.31)	(56.36)	(0.37)		(76.46)		
	[9.40%]	[13.52%]	[15.33%]		[10.38%]		
$H_0 : \beta_4^{Musculo} = \beta_4^{Other}$, p-value	<0.01	<0.01	<0.01		<0.01		
$H_0 : \beta_4^{Mental} = \beta_4^{Other}$, p-value	<0.01	<0.01	0.0720		<0.01		
Degree of disability							
Fully disabled	-10.95***	-1,656***	8.08***	4.88	1,037***	626	324,485
	(0.18)	(37.49)	(0.26)		(38.74)		
	[12.22%]	[17.05%]	[49.93%]		[51.03%]		
Partially disabled	-20.73***	-1,243***	4.00***	3.22	838***	674	172,101
	(0.35)	(57.40)	(0.41)		(99.65)		
	[26.24%]	[25.30%]	[6.06%]		[7.09%]		
$H_0 : \beta_4^{Full} = \beta_4^{Partial}$, p-value	<0.01	<0.01	<0.01		0.0670		
Partially disabled							
Not employed	-29.25***	-2,032***	10.90***	10.56	1,315***	647	44,087
	(0.76)	(156.4)	(1.04)		(274.3)		
	[32.48%]	[22.72%]	[53.93%]		[34.50%]		
Employed	-19.04***	-1,383***	-0.66	-0.48	548*	396	98,655
	(0.55)	(86.86)	(0.41)		(227)		
	[23.45%]	[18.83%]	[0.80%]		[2.94%]		
$H_0 : \beta_4^{Not} = \beta_4^{Employed}$, p-value	<0.01	<0.01	<0.01		<0.01		

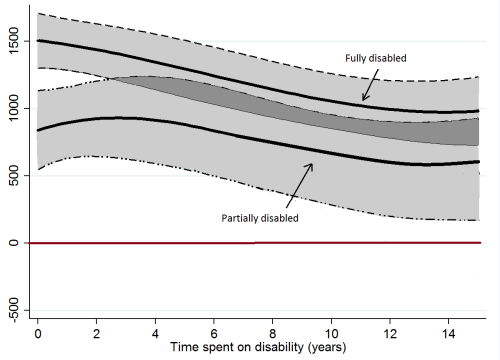
Notes: Group-specific least squares estimates of β_4 from (3) for the respective outcome. Standard errors adjusted for clustering at the individual level in parentheses. In square brackets is the estimated effect as a percentage of the predicted mean outcome under the counterfactual calculated as in footnote 16. p-values given for tests of equal effects across groups. Number of individuals is across all treatment and comparison groups. Number of observations is the number of individuals multiplied by 5. ppt = percentage points. *** indicates significance at the 1% level.

Figure 5: Earnings effect of reassessment by duration of claim - heterogeneity

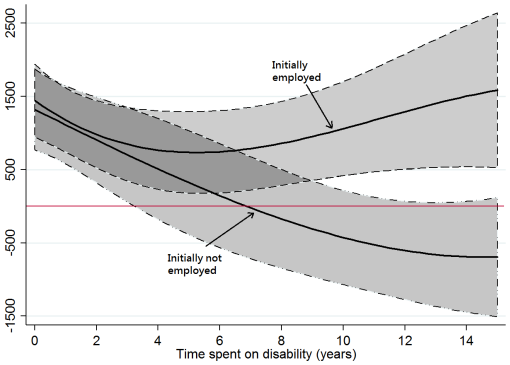
A: Cause of Disability



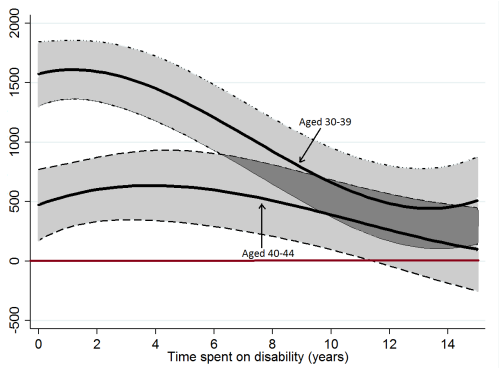
B: Degree of disability



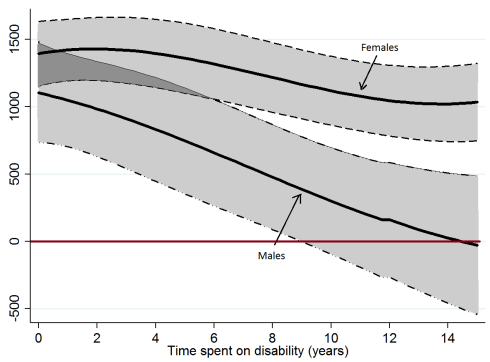
C: Employment (Partially disabled)



D: Age



E: Sex



Notes: Estimates of average effects of reassessment under stricter DI criteria by length of time on DI at entry to the sample. Derived from least squares estimates of $\beta_{4j}, j = 0, 1, 2, 3$ from (4) estimated separately for each sub-sample. Shading indicates 95% confidence interval computed from delta method standard errors adjusted for clustering at the individual level. Sample sizes are given in Tables 7 and 8.

5.5 Effects by age and sex

On average, we find that DI recipients had considerable earnings capacity they could be induced to utilize. This may be partly attributable to the age of those affected by the reform, who, at 30-44, are younger than those targeted by most other DI reforms that have been evaluated.[34] The top panel of Table 8 reveals that the work and earnings responses to reassessment are even stronger among claimants aged 30 to 39.[35] Their probability of employment increased by twice as much the respective increase among those aged 40-44. The employment effect relative to the predicted employment rate under the counterfactual is also twice as large for the younger group. In absolute terms and relative to the counterfactual mean, earnings also rise, on average, by twice as much in the younger group, which is able to recover 68% of the average reduction in DI income through increased labor market earnings compared with 55% of a smaller average loss for the older group. These results consistently indicate greater work and earnings capacity among the youngest DI recipients subjected to reassessment.[36]

The earnings response declines steeply with claim duration in the younger group but not the older group (Figure 5, Panels D). In fact, in the latter group there is even some sign of the inverted U-shape found by Moore (2015), which is consistent with a period on DI providing time to regain earnings capacity before the offsetting effect of labor market detachment begins to dominate.

In absolute terms, the employment response of female benefit recipients is almost twice

[34] Although the 1993 Dutch reform impacted all claimants below the age of 45, the research design employed by Borghans et al. (2014) identifies the effect at the margin of that age threshold only. Karlström et al. (2008) find no employment effect from withdrawal of laxer rules for those aged 60-64 to qualify for the Swedish DI program. At the slightly younger age of 55-56, Staubli (2011) finds a positive impact of reduced DI entitlement on employment in Austria. Moore (2015) finds that younger (30-39 vs 50-61) US SSDI recipients who qualified through an addiction had a larger employment response to the termination of their benefits. Three other papers also find a larger (in magnitude) elasticity of labor supply with respect to DI at younger ages (Koning and van Sonsbeek 2017; Kostol and Mogstad 2014; Von Wachter et al. 2011), although these studies do not identify from variation arising from benefit cuts.

[35] Splitting this group into those aged 30-34 and 35-39 reveals little further heterogeneity.

[36] Differences in the sex and disability composition of the age groups (see Appendix B Table B2), do not account for the difference in the earnings response by age. The younger group has a stronger earnings response irrespective of sex (Appendix C Table C5) and cause of disability (Appendix C, Table C6).

as large as that of male claimants (Table 8, bottom panel). Relative to the counterfactual (see figures in square brackets), the impact on the female employment rate is more than twice that on the male rate. Absolutely and especially relative to the counterfactual mean, the positive impact on market earnings of female recipients is considerably larger than the respective effect on male earnings. Women are able to increase their earnings to replace a larger fraction of their lost DI benefits. The average earnings effect is three quarters of the average reduction in DI income for female recipients, compared with three fifths for male recipients. Females were impacted more by the reform and, judging by their response to it, this targeting appears to have been justified.[37] The earnings effect declines steeply with length of time on DI among males but not females. Women who had been claiming DI for fifteen years were able to respond to reassessment by raising their earnings in excess of €1000 per year, on average.

[37]The age difference between male and female DI recipients (Appendix B Table B3) is not responsible for the different response by gender: within each age group, the employment and earnings effects are larger for females than for males (Appendix C Table C5).

Table 8: Effects of reassessment of DI recipients under more stringent rules by age and gender

	Disability Insurance		**Labor Market**				**No. individuals**
	Benefit Receipt (ppt)	Benefit Amount (€/year)	Employment (ppt)		Earnings (€/year)		
			Effect	Scaled effect	Effect	Scaled effect	
	(1)	(2)	(3)	(3)/|(2)| × 1000	(5)	(5)/|(2)| × 1000	(7)
Age							
30-39 years	-16.68***	-1,823***	8.55***	4.69	1,248***	685	330,042
	(0.23)	(36.09)	(0.27)		(51.01)		
	[20.19%]	[24.48%]	[25.17%]		[23.28%]		
40-44 years	-11.35***	-1,225***	4.30***	3.51	667***	544	363,412
	(0.22)	(39.47)	(0.27)		(53.30)		
	[12.73%]	[14.06%]	[12.27%]		[11.29%]		
$H_0 : \beta_4^{30-39} = \beta_4^{40-44}$, p-value	<0.01	<0.01	<0.01		<0.01		
Sex							
Males	-8.31***	-1,375***	4.21***	3.06	815***	593	244,076
	(0.25)	(49.79)	(0.32)		(72.91)		
	[9.76%]	[15.55%]	[10.80%]		[11.05%]		
Females	-18.32***	-1,769***	7.87***	4.45	1,338***	756	252,510
	(0.24)	(38.39)	(0.32)		(46.90)		
	[41.38%]	[42.43%]	[24.68%]		[31.71%]		
$H_0 : \beta_4^{Males} = \beta_4^{Females}$, p-value	<0.01	<0.01	<0.01		<0.01		

Notes as Table 7

6 Conclusion

To design policies that can relieve the fiscal strain caused by swelling disability insurance programs while protecting the wellbeing of benefit recipients with little latitude to raise their labor market earnings, it is critical to target benefit cuts on those with residual, unused earnings capacity. This paper is one of only a few that deliver evidence on the employment and earnings effects of targeted reductions in benefit entitlement in a stock of DI recipients. Using a 2004 reform that affected Dutch DI recipients aged 30-44, we find that reassessment, comprising a medical examination to review functional limitation and application of more stringent rules to calculate earnings capacity, reduced the probability of receiving DI benefits by 14.4 percentage points and reduced the benefit amount by around 20 percent, on average. These cuts induced a 6.7 percentage points increase in employment and an 18 percent average increase in earnings. A loss of €1000 in DI benefits was replaced, on average, by €636 of labor earnings.

These findings confirm that prior to the reform some recipients of disability benefits in the Netherlands had substantial capacity to substitute earnings for benefits. While the earnings response helped cushion the financial impact of the benefit cuts, the welfare of the affected recipients presumably fell. The consequences for social welfare are more difficult to determine. Since the earnings response could come from an income effect as well as being due to disqualified recipients no longer facing work disincentives that are built into DI rules, we cannot be sure that the reform raised efficiency by reducing the extent to which transfers crowd out earnings by distorting the return to work. It could also be that recipients impacted by the reform chose to enter work, or to work longer hours, simply because they had less non-labor income. Recent evidence suggests that much, perhaps even all, of the earnings crowd-out from US SSDI is an income effect (Gelber et al. 2017). In the Dutch context, it is likely that a greater part of the earnings response to the benefit cuts arising from reassessments of earnings capacity represents efficiency gains from reduced moral hazard. One reason is that the magnitude of the response makes it unlikely that is generated by

an income effect alone. A second reason is that even a decade into retrenchment of the Dutch DI program, the dependency rate was still higher than it is in the US. This reflected laxer qualification criteria and higher replacement rates that made the system vulnerable to moral hazard. The 2004 reform was intended to correct this by targeting benefit cuts on recipients with greater residual earnings capacity. The estimates suggest it was successful in this respect: upward revision of assessed earnings capacity induced behavioral changes to realize that capacity. A third reason is that we find a stronger earnings response from benefit recipients who had qualified through mental health and musculoskeletal conditions that are the most difficult to verify. While most claimants who qualify for DI through these conditions no doubt have earnings capacity that is insufficient to fully support themselves and their families, we find that, on average, they have greater potential to increase their earnings in response to benefit cuts than do claimants with more objectively defined conditions.

Our finding of large employment and earnings responses to the reassessment of benefit entitlement does not mean that all recipients of DI benefits in the Netherlands at the time of the reform, or even a majority of them, had abundant earnings capacity. While all benefit recipients younger than 45 had their entitlement reassessed under stricter criteria, benefits were left unchanged for almost three fifths of them. The strong earnings response is from a minority who lost benefits. Across-the-board cuts in DI benefits would reduce the gains from insurance and cause hardship among many benefit recipients who would be unable to raise their earnings to make up the loss of income. The policy difficulty lies in discriminating between recipients who do and do not have the capacity to replace benefits with earnings. The Dutch reform appears to have been successful in targeting cuts on those with the capacity to raise earnings. Our analysis reveals that they were disproportionately younger and female, and that earnings potential was also higher among those who had been claiming DI for a shorter period, were diagnosed with a more subjectively defined health condition and had been claiming partial disability benefits while not working.

These findings suggest that periodic, rigorous reassessment of the earnings capacity of the

stock of DI recipients can be effective in raising labor market activity by targeting benefit cuts on those most able to raise earnings in response. Such a policy, in contrast to the indiscriminate withdrawal of benefits that sometimes have only a muted impact on earnings (Koning and van Sonsbeek 2017) can help achieve the optimal balance between the provision of insurance and the preservation of work incentives.

References

Autor, D. H. (2015), The unsustainable rise of the disability rolls in the united states: Causes, consequences, and policy options, *in* J. K. Scholz, H. Moon and S.-H. Lee, eds, 'Social Policies in an Age of Austerity: A Comparative Analysis of the US and Korea', Edward Elgar, Northampton, MA, pp. 107–136.

Autor, D. H. and Duggan, M. G. (2006), 'The growth in the Social Security disability rolls: a fiscal crisis unfolding', *Journal of Economic perspectives* **20**(3), 71–96.

Autor, D. H., Maestas, N., Mullen, K. J. and Strand, A. (2015), Does delay cause decay? The effect of administrative decision time on the labor force participation and earnings of disability applicants, Working Paper 20840, National Bureau of Economic Research.

Autor, D., Kostol, A. R., Mogstad, M. and Setzler, B. (2017), Disability benefits, consumption insurance, and household labor supply, Working Paper 23466, National Bureau of Economic Research.

Bell, B., Blundell, R. and 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.

Blundell, R. and Costa Dias, M. (2002), 'Alternative approaches to evaluation in empirical microeconomics', *Portuguese Economic Journal* **1**(2), 91–115.

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

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

Bryngelson, A. (2009), 'Long-term sickness absence and social exclusion', *Scandinavian Journal of Public Health* **37**(8), 839–845.

Burkhauser, R. V., Daly, M. C., McVicar, D. and Wilkins, R. (2014), 'Disability benefit growth and disability reform in the us: lessons from other oecd nations', *IZA Journal of Labor Policy* **3**(1), 4.

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.

Cochran, W. G. (2007), *Sampling techniques*, John Wiley & Sons, New York.

De Jong, P., Lindeboom, M. and Van der Klaauw, B. (2011), 'Screening Disability Insurance applications', *Journal of the European Economic Association* **9**(1), 106–129.

French, E. and Song, J. (2014), 'The effect of Disability Insurance receipt on labor supply', *American Economic Journal: Economic Policy* **6**(2), 291–337.

Gelber, A., Moore, T. J. and Strand, A. (2017), 'The effect of Disability Insurance payments on beneficiaries' earnings', *American Economic Journal: Economic Policy* **9**(3), 229–61.

Karlström, A., Palme, M. and Svensson, I. (2008), 'The employment effect of stricter rules for eligibility for DI: Evidence from a natural experiment in sweden', *Journal of Public Economics* **92**(10-11), 2071–2082.

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

Koning, P. and van Sonsbeek, J.-M. (2017), 'Making disability work? The effects of financial incentives on partially disabled workers', *Labour Economics* **47**, 202–15.

Kostol, A. R. and Mogstad, M. (2014), 'How financial incentives induce Disability Insurance recipients to return to work', *American Economic Review* **104**(2), 624–55.

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

Ministry of Justice (2007), *Staatsblad van het Koninkrijk der Nederlanden*, Vol. 192, Sdu Uitgevers, The Hague.

Moore, T. J. (2015), 'The employment effects of terminating disability benefits', *Journal of Public Economics* **124**, 30–43.

OECD (2012), *Sick on the job? Myths and realities about mental health and work*, OECD Publishing, Paris.

Social Insurance Benefits Agency (UWV) (2009), *Rapportage afronding eenmalige herbeoordelingsoperatie*, Sdu Uitgevers, The Hague.

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

Svensson, T., Müssener, U. and 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.

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

Van Vuren, A. H. and Van Vuren, D. J. (2007), 'Financial incentives in disability insurance in the Netherlands', *De Economist* **155**(1), 73–98.

Vingård, E., Alexanderson, K. and Norlund, A. (2004), 'Consequences of being on sick leave', *Scandinavian Journal of Public Health* **32**, 207–215.

Von Wachter, T., Song, J. and Manchester, J. (2011), 'Trends in employment and earnings of allowed and rejected applicants to the Social Security Disability Insurance program', *American Economic Review* **101**(7), 3308–29.