Earnings responses to disability insurance stringency*

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

June 23, 2020

Abstract

Accurate assessment of earnings capacity is critical to the efficient operation of disability insurance (DI) programs. We use administrative data on the universe of Dutch DI recipients to estimate employment and earnings responses to reassessment of their earnings capacity under more stringent rules. We estimate that reassessment of recipients aged 30-44 removed 17 percent from the program and reduced benefit income 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 every €1000 reduction in DI benefit. This earnings response was strongest from those with more subjectively defined disabilities and a shorter claim duration, as well as younger and female recipients.

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

*We thank Wilbert van der Klaauw (Co-Editor), two referees, Courtney Coile, 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, sgarciamandico@gmail.com; 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 and Tinbergen Institute, 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. Evidence on the earnings response to reduced DI entitlement resulting from stricter assessment of the earnings capacity of benefit recipients can help determine whether the right balance has been struck.

This paper uses administrative data on the universe of Dutch DI recipients to estimate the impact on earnings, and employment, of reassessment of their earnings capacity under more stringent criteria that could result in benefits being terminated or cut substantially. If the reassessments were effective in identifying recipients with untapped earnings potential, then reduced benefits should have raised 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 to assess the effectiveness and targeting of the upward revisions made to earnings capacity. In doing so, we extend the evidence base on the labor supply effects of DI — the second largest item of social insurance expenditure in many countries — by adding to only a handful of studies that estimate effects of cutting the entitlement of current benefit recipients (Borghans et al. 2014; Deuchert and Eugster 2019; Deshpande 2016; Moore 2015).

We identify the effect of a 2004 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) between age groups to identify effects of more stringent criteria at *application* for DI (Karlström et al. 2008; Staubli 2011), we adjust for the age difference 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 difference between age groups in the outcome trends would have been the age group difference in the trends 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 over all DI recipients aged 30-44 targeted for reassessment, application of the more stringent rules reduced the amount of DI income received by 20 percent relative to what it would have been if there were no reform, raised employment by 20 percent and increased earnings by 18 percent; implying high elasticities of employment and earnings with respect to benefits. Receiving €1000 less from DI was compensated by earning €636 more in the labor market, on average. Apparently, some younger Dutch DI recipients had considerable untapped earnings potential that more stringent assessment of their benefit entitlement induced them to utilize.

The Netherlands provides an interesting context in which to assess the earnings potential of DI recipients. It is known for high DI dependency that reached 12% of the insured population at the beginning of the 1990s but also 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). 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 paring back of an initially generous program, 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 dependency was substantially higher than elsewhere.[1] Our estimate of the rate at which earnings replaced lost DI income is actually very close to that obtained by Borghans et al., indicating that even after a decade of retrenchment some DI recipients still had considerable unused earnings potential they

[1]Koning and Lindeboom (2015) argue that benefit cuts played a relatively minor role in reducing *entry* to DI in the Netherlands, while acknowledging the effect of the 1993 reform examined by Borghans et al. (2014) on *exit* from the program. They do not mention the 2004 reform we evaluate.

could call on to replace around two thirds of substantial reductions in benefits. However, it is important to emphasize that these were a minority of the stock of DI recipients. Most did not have their benefits reduced despite being subjected to reassessment of their earning capacity under more stringent criteria.

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 potential of successful applicants (Chen and Van der Klaauw 2008; Von Wachter et al. 2011).[2] Exploitation of plausibly exogenous variation in the award or appeal probability can eliminate upward bias in the estimated earnings potential *at the time of application* (Autor et al. 2017; French and Song 2014; Maestas et al. 2013). But this quasi-experimental strategy will still overestimate the average earnings potential of the stock of beneficiaries if skills and preferences for work deteriorate while on DI (Bryngelson 2009; Svensson et al. 2010; Vingård et al. 2004). 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 any earnings potential of benefit recipients.

There are only a few studies that, like this one and Borghans et al (2014), estimate labor supply responses of DI recipients to cuts made to their benefits. Moore (2015) finds that 22 percent of US Social Security Disability Insurance (SSDI) recipients entered employment after being removed from the program because they had (partly) qualified through an addictive disorder. Deshpande (2016) estimates that 18-year-olds removed from another US DI program following stricter medical review were able to increase earnings to an extent sufficient to replace only one third of the benefit income lost. Relative to the cuts made to benefits, these labor supply responses in the US are smaller than those of Dutch DI recipients that we and Borghans et al. estimate. This may be due to differences in the DI programs, but it could also reflect differences in the DI recipients studied. Those qualifying

[2] Chen and Van der Klaauw (2008) also obtain point estimates from a regression discontinuity in DI entitlement based on age.

through an addiction were only two percent of the stock of SSDI recipients, and their work preferences and capacities may have been quite distinct. At the age of 18, the DI recipients studied by Deshpande lacked the labor market experience that may have conditioned their labor supply response to benefit cuts. We estimate responses of all recipients aged 30-44, who comprise more than a third of the stock of DI recipients in the Netherlands, where, as in other countries, the DI roll is becoming younger.

Besides being one of the few studies to estimate earnings responses to targeted reductions in the benefit entitlement of DI recipients, this paper adds to the meager evidence on whether and how these responses vary with time spent on DI (Autor et al. 2015; Gelber et al. 2017; Moore 2015). Using claim durations of up to 15 years, which is substantially longer than other studies, we find that reassessment induced a smaller earnings response from those who had been claiming for longer. Interestingly, the earnings response of partially disabled recipients who were working at the time of reassessment did not decline with claim duration.

We find that DI recipients who qualified through more subjectively defined health problems — mental health and musculoskeletal conditions — experienced the most aggressive cuts in benefits, indicating the greatest upward revisions in assessed earnings capacity, and were able to increase earnings to replace larger fractions of these cuts. This is consistent with the argument that loosening of the criteria for DI entitlement from precisely defined medical diagnoses to the more nebulous concept of *work capacity* lengthened DI rolls by opening the door to claims based on difficult-to-verify health problems (Autor 2015; Autor and Duggan 2006).[3]

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

[3]In 2012 across all OECD countries, mental health disorders were cited as the cause of one half of ongoing DI claims (OECD 2012). Musculoskeletal problems are typically the second most common reason given for a DI claim. Studies based on comparisons between accepted and rejected DI applicants in the US produce contradictory evidence on whether claimants citing more subjective health problems have greater earnings capacity (French and Song 2014; Maestas et al. 2013; Von Wachter et al. 2011). Moore (2015) finds that among SSDI recipients who had partly qualified through an addiction, those with a primary diagnosis of a mental health or a musculoskeletal condition were more likely to work after their benefits were terminated.

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 the relationship between earnings responses and claim duration, and then heterogeneity analyses. The final section concludes.

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

[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%. Those less than fully disabled receive this earnings-related benefit for a limited period (6 years max). See Appendix A.1 for further details.

up to the maximum earnings consistent with their assessed degree of disability. Earning 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: reassessment under more stringent rules

From October 2004, the stock of DI 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. The criteria used in this part were the same as previously, and so it could result in revision of the recipient's assessed functional limitations only if their health condition was observed to have changed. Descriptive analysis presented in Appendix A.2 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 (see Appendix A.2 for details). 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 could be cut or terminated. This 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).

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

[6]Plans for the reform were announced in May 2003 and the reform was legislated in April 2004, with the intention 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 2004. 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.

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).[7] Consequently, we restrict attention to benefit recipients aged 30-44 on July 1, 2004.

Among those DI recipients, about a quarter (24.4%) were reassessed as having a degree of disability below the 15% minimum threshold and had their entitlement withdrawn completely. Almost half (47.9%) of those initially with the lowest degree of disability [15%, 25%) were disqualified from receiving any benefit. 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 aged 30-44 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. The initially fully disabled were least affected: 72% continued to received the same amount of benefit.[8] 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.[9]

The consequences of the reform for benefit entitlement were clearly heterogeneous. Greater downward revision to the degree of disability resulted in a larger reduction 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 eligible for reassessment, a majority of whom

[7]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.4). 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.

[8]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.

[9]See Appendix A.2 Table A.1 for detailed analysis of the changes in degree of disability resulting from the reassessments.

experienced no change in their benefit entitlement. The average effect will be much smaller than the average reduction in benefits paid to those whose degree of disability was reduced as a result of reassessment.

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, more than four fifths (81%) had been undertaken by the end of 2006 and they were all but completed (99.9%) by the end of 2008 (see Appendix A.3 Table A.2). Around 14% of those who had been claiming DI in January 2004 and who were eligible by age for reassessment — the two characteristics that define our treatment group — left the program before there was an opportunity to reassess them. Since they may have exited in response to the prospect of reassessment, we include these individuals in the treatment group used to estimate effects of the reform.

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 with the outcome of reassessment in a way that suggests recipients who the agency expected would experience larger benefit cuts were called earlier (see Appendix A.3). For this reason, we do not attempt to exploit variation in the timing of reassessment for identification.

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 of the reform, comprising reassessment of the stock of younger DI recipients under more stringent rules, on benefit receipt and labor supply. To estimate average effects on recipients aged 30-44, we need a comparison group that allows credible identification of the average outcomes 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 being targeted for 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 (Appendix Table A.2). We test robustness to using monthly data, which avoids this inaccuracy, in section 5.3.[10] Let *t=4* be four years later in 2008 when the reassessments were completed (but for a negligible $< 0.001\%$). Then, $Y_{i4} = D_i Y_{i4}^1 + (1 - D_i) Y_{i4}^0$, where $D_i = 1$ if i has been targeted for reassessment and is 0 otherwise. We wish to estimate the average effect of the reform on those targeted for reassessment: $ATET = \mathbb{E}\left[Y_{i4}^1 - Y_{i4}^0 \mid D_i = 1\right]$.[11]

[10] 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.

[11] We take 2008 as the endpoint because of a data constraint explained below. This risks not capturing the full effect on the 2.9% who were reassessed during 2008. Since most of them were reassessed at the beginning of 2008, and also because they had longer to prepare for reassessment and so may have responded more quickly, any downward bias should be modest. We define treatment to include those who left DI before they could be reassessed since leaving DI may have been a response to the prospect of reassessment. It should be kept in mind that the target group had warning of this prospect and this may have influenced the effect of the reform.

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.[12] This is likely to be problematic since older DI beneficiaries have a lower probability of returning to work and recovering their earnings than younger recipients, even 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 ($AGE_i = 0$). Effectively, we subtract the age-differential trend in the *non-reform period* from the age group DID 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 assuming that the change in earnings in the 30-44 age group would have been the same in the two periods if there had

[12] 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.

been no reform in the later period. 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:

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

We assess the plausibility of this assumption in section 4.3 by comparing age differences in trends across periods in which there was no reform. If the 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 the reform on those targeted for reassessment is then given by the DADID:

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

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.[13] But the pre-reform trends presented in section 4.3 do not reveal patterns consistent with anticipation. If the effect of the reform were to have spilled over to reduce the labor market activity of the older group, possibly through intensified job competition from the younger, targeted group or because implementation of the reform diverted the benefits agency from conducting periodic, standard reassessments of the older group, then the magnitudes of

[13] The planned reform was initially announced in May 2003, and so it is possible that it was anticipated by those aged 30-44 in our non-reform period cohort, as well as those of the same age in the reform period cohort. If there were behavioral responses to any such anticipation already in 2003 in either or both cohorts aged 30-44, then our DADID estimates of effect magnitudes will be downwardly biased.

our estimates would be upwardly biased. However, the risk of spillover bias is substantially reduced by the very low rate of exit of the older group from DI (around 5%) even in normal times. A six year age gap between the two groups further reduces the risk. If the bias were present, it would be evident from the older group's outcome trends in the reform period diverging from those in the non-reform period, which is not the case (see Appendix B Figure B1). While we cannot rule out spillover bias entirely, the context and descriptives suggest that it is unlikely to be anything other than negligible. In section 5.2, we further assess the credibility of the strategy by checking that it gives a zero "effect" on the earnings of individuals who were not DI recipients and so were not exposed to the reform.

3.2 Estimation

To estimate the effects, we pool two 5-year balanced panels of DI recipients from the reform period (2004-2008) and the non-reform period (1999-2003). 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,

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

where $YEAR_t, t = 0, 1, ..4$ 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, $YEAR_1 = 1$ & $PERIOD_i = 1$ indicates 2005 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 reassessments 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 (Appendix A.3 Table A.2). We focus on the estimate of β_4, which corresponds to the $ATET$ of the reassessment reform. Note that we are estimating the effect of the reassessment reform, 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.[14] 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.[15]

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.

[14]The classification uses the most aggregated level of the International Classification of Diseases version 9.

[15]The estimated effect on employment is highly robust to defining employment as being in paid work for a minimum of a week, a month and a quarter, as opposed to a day (see Appendix C Table C7).

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.[16] 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. The group includes 22,380 individuals who left DI before the agency managed to reassess their eligibility. Since these exits may have been in anticipation of the outcome of reassessment, these individuals can be considered to have been exposed to the reform and are appropriately part of the treatment group.

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 spanning the years 1999-2003 with that constructed for

[16]DADID estimates of the effects of the reform on individuals aged 45-49 are given in Appendix C.1, Table C1. As expected given this group's diluted exposure to the reform, the effects on the receipt of DI, employment and earnings are all the same sign but considerably smaller in magnitude compared with those for the 30-44 age group presented in Table 2. The effect on the benefit amount received by those aged 45-49 is positive (in 2008). This surprising result is likely due to compensation paid to recipients who had their benefits cut temporarily (see Appendix C.1).

the reform period, 2004-2008.

Table 1 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 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.

Table 1: 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 (years)	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 individuals	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).

4.3 Trends

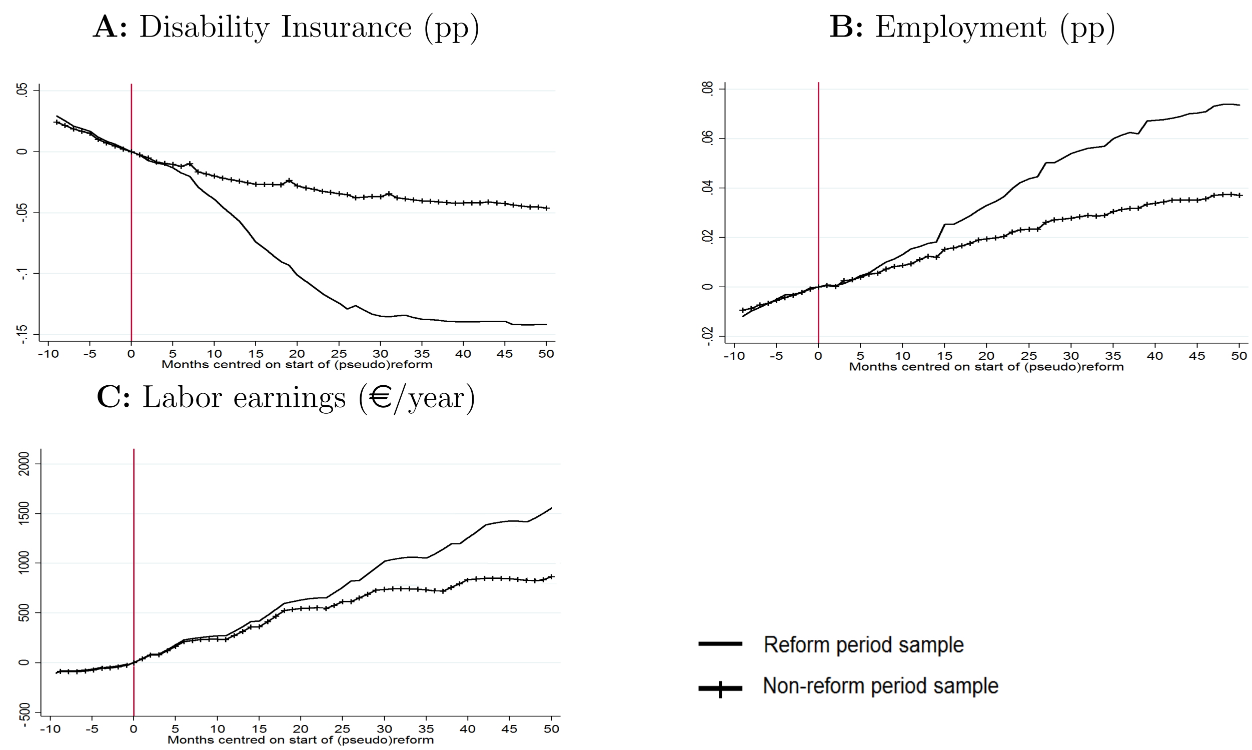
Figure 1 shows difference-in-differences in receipt of any DI benefits, employment and labor earnings between the two age groups within each period.[17] These figures are drawn using monthly data to allow more detailed assessment of the evolution of the trends before and after the start of the reassessments. Each line traces the age group difference (30-44 years - 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.

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

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

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

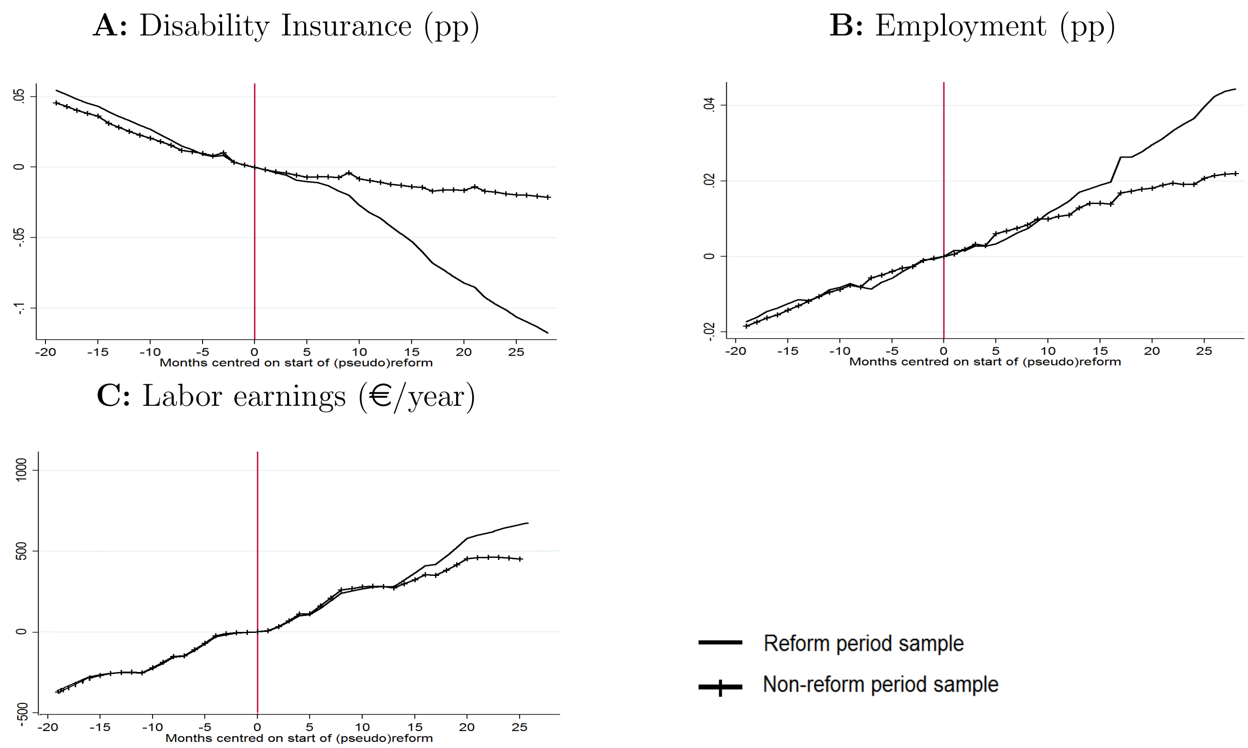


Note: Reform period (Jan. 2004-Dec. 2008) sample consists of individuals aged 30-44 & 50-53 on July 1, 2004 who were claiming DI in January 2004. Non-reform period (Jan. 1999-Dec. 2003) sample consists of individuals aged 30-44 &50-53 on July 1, 1999 who were claiming DI in January 1999. Month 0 is October 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 1. pp = percentage points.

credible, we show in Figure 2 two different cohorts of DI recipients traced over 21 months prior to the start of reassessments in the reform period.[18] The age differentials in the outcome trends do not diverge markedly between the two cohorts over this extended time span before reassessments started. 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 the case in the earlier period sample. While 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.

[18]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: 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.

5 Results

5.1 Main estimates

Column (1) of Table 2 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.[19] The middle column gives the treatment group's predicted mean outcome in 2008 under the counterfactual of no reform, i.e. $\frac{1}{n_T}\sum_i (AGE_i \times PERIOD_i \times YEAR_4)\,\hat{Y}_{it} - \hat{\beta}_4$, where $\hat{Y}_{it}$ is the predicted outcome from (3) and n_T is the number of individuals in the treatment group. Column (3) gives effects on labor market outcomes and other social transfer income scaled by the estimated effect on 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.[20]

We estimate that reassessment reduced the probability of remaining on DI in 2008 by 14.4 percentage points.[21] This includes the direct effect of claims terminated through application of the stricter rules as well as any indirect effect that may arise through reduced benefits inducing some to leave DI. 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

[19]The estimated effects in all the post-reform years are given in Appendix C.2 Table C2. The effects increase in magnitude with time since the start of the reform period, which reflects the growing number of recipients who are reassessed.

[20]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 health had deteriorated sufficiently to offset the effect of increased stringency. Hence, monotonicity does not hold.

[21]This is somewhat larger than an estimate obtained by taking the difference between the reform period and non-reform period difference-in-differences at the extreme right of panel A of Figure 1. Employment and earnings effects estimates in Table 2 are also a little larger than those inferred from panels B and C respectively of Figure 1. The reason is that Figure 1 is drawn using monthly data, while the Table 2 estimates are obtained from yearly data. Robustness to using monthly data is assessed in Table 3, Panel D.

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.[22] Given that the degree of disability did not change as a result of reassessment for a majority and it even increased for a few (Appendix A.2 Table A.1), 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.[23] If the magnitudes of the two effects were equal, 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 degree of disability was raised, then the average effect on the 34% whose benefits were cut would still be a substantial €4549.[24] This is 54% of the mean benefit income received by the treatment group prior to the reform.

[22]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.

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

[24]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 2: Effects of reassessment of DI recipients under more stringent rules

	Effect	Predicted mean if no reform	Effect scaled by benefit reduction (in €'000s/year)
	(1)	(2)	(3)
Disability Insurance			
Benefit Receipt (pp)	-14.40***	84.52	NA
	(0.20)		
Benefit Amount (€/year)	-1,565***	7,906	NA
	(47.60)		
Labor Market			
Employment (pp)	6.68***	33.83	4.27
	(0.25)		
Days worked (year)	17.03***	76.26	10.88
	(0.68)		
Earnings (€/year)	995***	5,507	635.8
	(43.19)		
Other social transfers			
Benefit amount (€/year)	376***	877	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 predicted mean outcome of 30-44 age group in 2008 under counterfactual of no reform, i.e. $\frac{1}{n_T}\sum_i (AGE_i \times PERIOD_i \times YEAR_4)\,\hat{Y}_{it} - \hat{\beta}_4$, where $\hat{Y}_{it}$ is the predicted outcome from (3) and n_T is the number of individuals in the treatment group. Columns (3) gives 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 1. pp = percentage points. *** indicates significance at the 1% level.

Having established that the reform reduced DI entitlement, 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 corresponds to a 4.3 point rise in employment set against a €1000 loss in annual income received from DI (Table 2).[25]

Borghans et al. (2014) estimate that a 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) finds that 22 percent of US SSDI recipients whose benefits were terminated entered employment. Relative to a 100 percent loss of benefit entitlement, this is a much smaller employment response than we find.[26]

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 each €1000 reduction in DI benefit received, €636 could be regained through labor market earnings.[27] This is very close to the €618 estimated by Borghans et al.. The recovery of

[25]We estimate that the probability of working and not claiming DI was increased by 8.5 points (SE=0.18, p-value<0.01). Given this is larger than the effect on the unconditional probability of employment, reassessment reduced the likelihood of claiming DI and working (by 1.8 points). This is likely due to initially partially disabled working claimants being forced or induced to leave the program.

[26]Besides the addictive behavior of those targeted by the reform evaluated by Moore, the difference could partly arise from incentives for disqualified US claimants to stay out of work in order to strengthen their case at reapplication. There is no such incentive in the Dutch system.

[27]Bearing in mind that the estimated reduction in DI benefits arises not only directly from cuts imposed

two-thirds of lost benefit income through increased earnings is double the rate managed by the US 18-year-olds who lost their DI entitlement studied by Deshpande (2016). In the Netherlands, even after the 1993 reduction in entitlement, some DI recipients subjected to reassessment in 2004 still had considerable earnings potential they could be induced to utilize 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 1).

It bears emphasis that these are average effects and reassessment resulted in the reduction or termination of benefits for a little more than one third of recipients (Appendix A.2 Table A.1). 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.[28] 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 reduced DI entitlement increased the amount received from other social

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.

[28]In addition to the reasons given in footnote 24 for expecting the magnitude of any effect on the benefit entitlement of recipients whose degree of disability (DD) was increased 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 due to a benefit reduction (with constant health). Using the formula given in footnote 23, 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.

transfers by €376, on average (Table 2).[29] This is 24% of the average reduction in income received from DI. The respective 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 social transfer income gives a total of €1371, which is about 88% of the estimated average reduction in payments received from DI.

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 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. 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. We use a random 50% sample of the 6.7 million individuals available for analysis.

We get precisely estimated zero "effects" on earnings and days worked (see Appendix C.3 Table C4). There is a small, but statistically significant, negative "effect" on employment.[30]

[29] Around half of the spillover to other programs was to unemployment insurance (UI) (Appendix C.2 Table C3). Those deemed ineligible for DI were automatically transferred to UI if they had made sufficient social insurance contributions prior to entering DI.

[30] The direction of this effect may seem puzzling given that macroeconomic conditions were better in 2004-2008 than they were in 1999-2003. The explanation is that it is an age difference in the period effect

Significance may simply be attributable to the huge sample. The point 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 the other two labor market outcomes.

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 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. It is credited with substantial reductions in the rate of DI inflow (De Jong et al. 2011; Koning and Lindeboom 2015; Van Sonsbeek and Gradus 2012). Any impact on the exit rate, as well as on the employment and earnings of those already receiving DI, would be indirect, and would not necessarily differ by age. Nonetheless, we test whether the GP may be confounding our estimates by dropping all DI recipients who had been claiming for 12 months or less at the time of selection into the reform period — who were potentially impacted by the GP — and drop the equivalent recipients from the non-reform period panel.[31]

on the trend, not simply a period effect. See Appendix C.3 for further explanation.

[31]The GP reform 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 estimated effects on DI benefit amount and employment given in panel B of Table 3 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 resulted in a loss of benefit income of €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 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 3 provides estimates 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 (Appendix A.2 Table A.2). 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.[32]

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.

[32]See footnote 10 for the reasons monthly data are not used to obtain the main estimates. 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

Table 3: Robustness to alternative sample selections and use of monthly data

	Disability Insurance		Labor Market			
	Benefit Receipt (pp)	**Benefit Amount** (€/year)	**Employment (pp)** Effect	Scaled effect	**Earnings (€/year)** Effect	Scaled effect
	(1)	(2)	(3)	$(3)/\lvert(2)\rvert \times 1000$	(5)	$(5)/\lvert(2)\rvert \times 1000$
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						
	-11.57***	-1,521***	4.17***	3.73	784***	515
	(0.37)	(65.2)	(0.46)		(93.1)	

Notes: Panel A reproduces the main estimates from Table 2 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 2.

The estimates obtained (Table 3, Panel D) are somewhat smaller than the main estimates, which is inconsistent with the latter not capturing the full effects of the reform. The effect on receipt of DI is 2.8 percentage points smaller than the main estimate and employment effect is 1.5 points smaller, while 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 reductions in DI benefits, and both employment and earnings increased markedly to compensate.

5.4 Effects by duration of claim

We now turn to the question of whether earnings potential and preferences for work diminish while claiming DI. 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. This is done by interacting a third order polynomial of that duration with the product of the age group, period and year indicators that identifies exposure to the reform (see Appendix C.4 for details).[33] To avoid heterogeneity by correlated characteristics confounding differences in earnings responses by claim duration, we also estimate the extended model that includes the reform-duration interaction for separate demographic and disability groups.[34]

Using the full sample without stratification, estimates from the extended model imply that the effects do indeed vary with the duration of a claim.[35] This variation is traced in Figure 3 that shows estimates up to a duration of 15 years, which corresponds to the 80^{th} percentile in the treatment group. The effects 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 at the end of this period.

[33] 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 coefficients on the interactions.

[34] 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 5.6.

[35] For each of the outcomes, the coefficients of the interactions between the reform indicator and the third order polynomial in duration are separately and jointly statistically significant (Appendix C.4, Table C6).

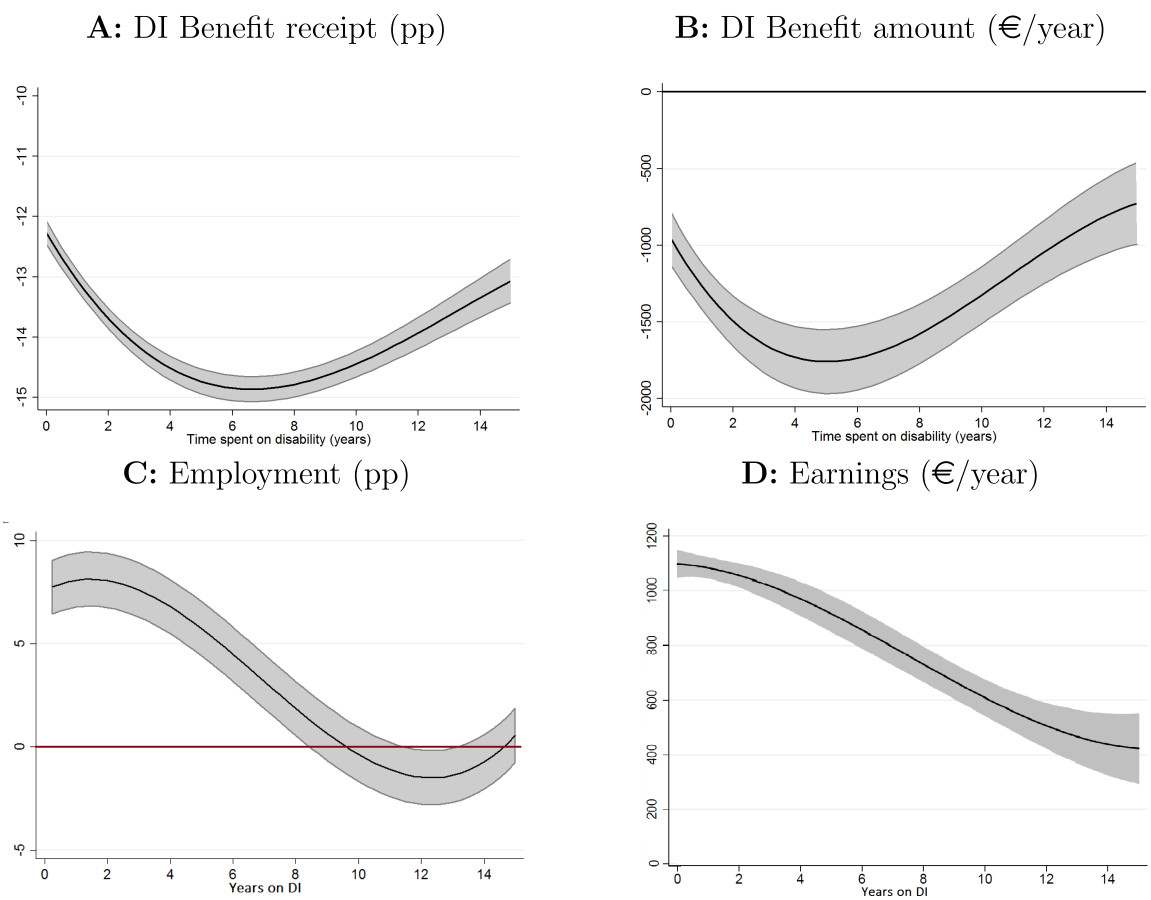
reform increases up to about 6 years. Thereafter, both effects fall in magnitude. The initially increasing effect on the benefit amount implies that medical reexamination and application of more stringent entitlement rules resulted in greater upward revision to assessed earnings capacity as claim duration lengthened up to 6 years. The decreasing revision from then on is consistent with deterioration in earnings capacity as the claim lengthens 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, peaking after about three years, that Moore (2015) finds with US data. The author suggests this could reflect a rehabilitation effect — time on DI initially provides an opportunity to recover health and work capacity — that is eventually dominated by increasing labor market detachment as the duration of the claim lengthens. Besides the obvious institutional differences, the discrepancy from the pattern we find may be partly due to the different nature of the reforms examined.[36] 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. However, even after scaling the employment effect by the benefit effect, the increase in employment relative to a €1000 reduction in benefit income still declines steadily with lengthening claim duration (see Appendix C.4 Figure C1).

The absolute 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. Scaling by the benefit effect reveals a decreasing fraction of lost benefits are replaced by earnings as claim duration lengthens (Appendix C.4 Figure C1). Those

[36]Two other US studies do not find the same pattern as Moore (2015). Autor et al. (2015) find that employment and earnings fall even as the time waiting for a SSDI application to be decided lengthens, while Gelber et al. (2017) find that labor supply responses to this program vary little with claim duration.

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 $\beta_{4j}, j = 0, 1, 2, 3$ in model (1) that appears in Appendix C.4. Estimates of these parameters are given in Appendix C.4 Table C6). Shading indicates 95% confidence interval computed from delta method standard errors adjusted for clustering at the individual level. pp = percentage points. Sample size is 496,586 individuals and 2,482,930 observations.

with short claim durations (less than a year) overcompensated through earnings for reduced benefits. But those who had been claiming for 8 years or more were able to recover less than half of more modest benefit reductions. Recipients who had been claiming for many years had their assessed earnings capacity revised upward to a lesser degree and they were less able, or possibly less willing, to increase their earnings to realise this more modest adjustment.

5.5 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 potential, 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—and estimate the regression model (3) separately for each sub-sample. The estimates given in the top panel of Table 4 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.[37] 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 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. Alternatively, the low absorption rate reflects greater distaste for work among those who had qualified for DI through these two less easily verifiable types of conditions. Unfortunately, we cannot assess the relative veracity of these two explanations.

In absolute terms, the average earnings response is about twice as large for those with

[37] These differences, like all other heterogeneous effects referred to in the text, are significant. For all heterogeneous effects on days worked and on income from other social transfers, see Appendix C.6 Table C9.

Table 4: Effects of reassessment of DI recipients under more stringent rules by cause and degree of disability

	Disability Insurance		Labor Market		No. individuals
	Benefit Receipt (pp)	Benefit Amount (€/year)	Employment (pp)	Earnings (€/year)	
Cause of disability					
Musculoskeletal	-19.81***	-2,015***	7.82***	1,221***	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***	1,156***	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***	620***	174,816
	(0.31)	(56.36)	(0.37)	(76.46)	
	[9.40%]	[13.52%]	[15.33%]	[10.38%]	
Degree of disability					
Fully disabled	-10.95***	-1,656***	8.08***	1,037***	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***	838***	172,101
	(0.35)	(57.40)	(0.41)	(99.65)	
	[26.24%]	[25.30%]	[6.06%]	[7.09%]	
Partially disabled					
Not employed	-29.25***	-2,032***	10.90***	1,315***	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	548*	98,655
	(0.55)	(86.86)	(0.41)	(227)	
	[23.45%]	[18.83%]	[0.80%]	[2.94%]	

Notes: Group-specific least squares estimates of β_4 from (3) for the respective outcome. Obtained by stratification by group. 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 i.e. $\frac{1}{n_T}\sum_i (AGE_i \times PERIOD_i \times YEAR_4)\,\hat{Y}_{it} - \hat{\beta}_4$, where $\hat{Y}_{it}$ is the predicted outcome from (3) and n_T is the number of individuals in the treatment (sub-)group. Tests of equality of the β_4 parameters between all groups compared (musculoskeletal vs other, mental vs other, full vs partially disabled etc. give p-values <0.01 for all outcomes and comparisons with two exceptions: employment, mental vs other p=0.072; earnings, full vs partial, p=0.067. Number of individuals is across all treatment and comparison groups. Number of observations is the number of individuals multiplied by 5. pp = percentage points. *** indicates significance at the 1% level.

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 received. Overall, benefit recipients with mental health and musculoskeletal problems appear to have had 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.[38] 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 (see Appendix C.6 Figure C2). This is consistent with work capacity and/or preferences deteriorating during the time spent on DI only among those who qualify through more difficult to verify health problems.

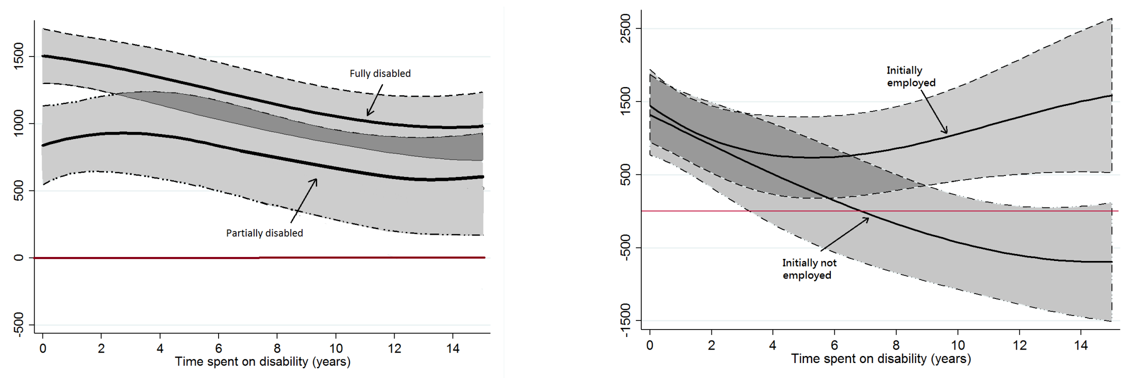
Among the initially partially disabled, reassessment reduced the probability of remaining on DI by twice as much as it did among the fully disabled (Table 4, middle panel). Despite this, the employment effect on the partially disabled is only half as large as that on the fully disabled. This is because 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). The bottom panel of the table reveals that reassessment raised the employment of the partially disabled who were not initially working by almost as much as it raised the employment of the fully disabled. The fully disabled increased their earnings by more than the partially disabled both absolutely and relative to the counterfactual mean. But this is because there was a greater fall in the benefit income received by the fully disabled. For a €1000 reduction in the benefit received, the increase in earnings is similar (€626-674). Even those who had been

[38] The earnings responses are largest for those with mental and musculoskeletal conditions within all sex, age and degree of disability groups (Appendix C.6 Tables C13 and C12).

classified as fully disabled were able to increase earnings to an extent sufficient to replace more than three fifths of the benefit income lost.

Figure 4: Earnings effect of reassessment by claim duration – heterogeneity

A: Degree of disability **B:** Employment of partially disabled)



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 (1) (see Appendix C) 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 Table 4.

The earnings effect of reassessment declines more steeply with time on DI among fully disabled recipients than it does among the partially disabled (Figure 4, Panel A).[39] But splitting the partially disabled by initial employment status reveals that the relationship mainly differs with the latter characteristic (panel B). 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.[40] This is consistent with partial disability combined with employment being effective in maintaining earnings capacity and work preferences as the time claiming DI lengthens, although it does not demonstrate this effect.

[39]Given the effect decreases with claim duration among the fully disabled, the presence of this relationship in the full sample is not simply due to recipients with longer durations being more disabled.

[40]This is not because the relationship between benefit cuts and claim duration differs between the initially working and non-working partially disabled (see Appendix C.6 Figure C3 Panel B and Figure C4 Panel B).

5.6 Effects by age and sex

On average, we find that DI recipients had considerable earnings potential 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.[41] The top panel of Table 5 reveals that the work and earnings responses to reassessment are even stronger among recipients aged 30 to 39.[42] Their probability of employment increased by twice as much as 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 (see estimates in square brackets). 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 realizable work and earnings potential among the youngest DI recipients subjected to reassessment.[43]

In absolute terms, the employment response of female benefit recipients is almost twice as large as that of male claimants (Table 5, bottom panel). Relative to the counterfactual, 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

[41] Borghans et al. (2014) examine a 1993 Dutch DI reform that impacted all recipients below the age of 45 but estimate the effect only at the upper margin of that age threshold. 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. Staubli (2011) finds a positive impact of reduced DI entitlement on the employment of 55-56 year-olds 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 greater 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).

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

[43] The younger group has a stronger earnings response irrespective of sex and cause of disability (Appendix C.6, Tables C11) and C12). On average, the younger group was reassessed five months earlier than the older group. If having more time to prepare for reassessment increases the labor market activity response to it, then we are underestimating the ceteris paribus difference in this response by age.

Table 5: Effects of reassessment of DI recipients under more stringent rules by age and gender

	Disability Insurance		**Labor Market**		**No. individuals**
	Benefit Receipt (pp)	Benefit Amount (€/year)	Employment (pp)	Earnings (€/year)	
Age					
30-39 years	-16.68***	-1,823***	8.55***	1,248***	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***	667***	363,412
	(0.22)	(39.47)	(0.27)	(53.30)	
	[12.73%]	[14.06%]	[12.27%]	[11.29%]	
Sex					
Males	-8.31***	-1,375***	4.21***	815***	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***	1,338***	252,510
	(0.24)	(38.39)	(0.32)	(46.90)	
	[41.38%]	[42.43%]	[24.68%]	[31.71%]	

Notes: Tests of equality of the effects (β_4 parameters) between age groups and between genders give p-values <0.01 for all outcomes and comparisons. Otherwise, notes as Table 4.

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.[44]

The earnings response declines steeply with claim duration for males and the younger age group but not for females and the older group (see Appendix C.6 Figure C2).[45]

6 Conclusion

Relieving 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 requires targeting benefit cuts on those with residual, unused earnings potential. This paper is one of only a few that deliver evidence on the employment and earnings effects of targeted

[44]Within each age group, the employment and earnings effects are larger for females than for males (Appendix C.6 Table C11).

[45]The earnings effect scaled by the effect on benefit income declines with claim duration for both genders and both age groups (Appendix C.6 Figure C4).

reductions in the benefit entitlement of a stock of DI recipients—those aged 30-44 in the Netherlands. Reassessment through a medical examination to review functional limitation and application of more stringent rules to assess earnings capacity reduced the probability of being on DI by 14.4 percentage points and the amount of benefit received by around 20 percent, on average. In response, employment rose by 6.7 percentage points and earnings increased by 18 percent. For each €1000 reduction in DI benefit there was a €636 increase in earnings, on average.

While these estimates indicate that some Dutch DI recipients had substantial potential to substitute earnings for benefits, they were not a majority. All DI recipients younger than 45 had their entitlement reassessed under stricter criteria, and yet benefits were left unchanged for almost three fifths of them. The strong earnings response is from a minority who fully or partially lost benefits. Across-the-board cuts would have reduced the gains from insurance and may have caused hardship among many benefit recipients who would have been unable to raise their earnings to make up the loss of benefit income.

The Dutch reform appears to have been successful in targeting cuts on those with the potential and inclination to respond by raising their earnings. Our analysis reveals that they were disproportionately younger, female and diagnosed with a more subjectively defined health condition, and that earnings potential was also higher among those who had been claiming DI for a shorter period. The greater responsiveness found at younger ages, together with the fact that the group targeted by the reform was younger (30-44) than those studied in many papers that have found labor outcomes to be less responsive to DI (Chen and Van der Klaauw 2008; Karlström et al. 2008; Staubli 2011), implies some limit on the extent to which our results generalize. But given the falling average age of the stock of DI recipients in many countries, estimates of the earnings potential of younger recipients are of increasing policy relevance. Another characteristic that distinguishes the DI recipients examined in this paper from those studied in many others is the high proportion ($\approx 36\%$) claiming partial disability. The opportunity afforded to this group to work while claiming DI may be expected to leave it better placed to substitute earnings for lost benefit income. If so, this could also explain why we obtain relatively large estimates of the responsiveness of labor outcomes to DI benefits. It turns out that while the partially disabled do manage to increase their earnings to recover a slightly larger fraction of the reduction in benefit income, the difference is generated by those who were not working initially. However, there is some evidence at least consistent with partial disability combined with work reducing the extent to which earnings potential declines as more time is spent on DI.

Our findings suggest that periodic, rigorous reassessment of the earnings capacity of the stock of DI recipients could be effective in raising labor market activity by targeting benefit cuts on those most able to respond by raising earnings. Such a policy, in contrast to the indiscriminate withdrawal of benefits that sometimes has only a muted impact on earnings (Koning and van Sonsbeek 2017), can possibly help achieve the optimal balance between the provision of insurance and the preservation of work incentives. That said, we cannot be sure that the reform examined improved efficiency. One reason is that while targeted cuts can be more effective in inducing earnings responses, they are also costlier to implement than across-the-board cuts that do not require manpower intensive case-by-case assessment. To implement the reform, the benefits agency received an additional budget of €190m to cover the hiring of new staff and the reintegration of recipients from whom benefits were withdrawn (Minister of Social Affairs 2009). In addition, it is estimated that expenditure on other social insurance and assistance programs rose by €550m as DI recipients sought to replace their lost DI benefits with other transfers (ibid). Subtracting these two costs of the reform from an estimated gross reduction in DI benefits of €1.1bn gives a net saving to the benefits agency of €375m (ibid). From a fiscal perspective, the targeted cuts were beneficial. However, they presumably reduced the welfare of DI recipients who had to increase labor supply effort to protect living standards. If this earnings response largely consisted of an income effect of lower non-labor income, as opposed to a price effect of the increased return on work, then it would mostly signal change in the distribution of welfare resulting from redistribution of income, rather than a net gain in welfare through the correction of incentives. The income effect of DI has been found to be much stronger than its substitution effect in both the US (Gelber et al. 2017) and Switzerland (Deuchert and Eugster 2019). Unfortunately, it is not possible to disentangle these two effects of the reform examined in this paper.

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.

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.

Deshpande, M. (2016), 'Does Welfare Inhibit Success? the Long-Term Effects of Removing Low-Income Youth from the Disability Rolls', *American Economic Review* **106**(11), 3300–3330.

Deuchert, E. and Eugster, B. (2019), 'Income and substitution effects of a disability insurance refrom', *Journal of Public Economics* **170**, 1–14.

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.

Minister of Social Affairs (2009), *Structuur van de uitvoering werk en inkomen (SUWI); Brief minister ter voldoening aan toezeggingen die hij heeft gedaan tijdens het Algemeen Overleg van 1 juli 2009 over Arbeidsongeschiktheid, Kamerstuk 26 448 nr. 411*, Tweede Kamer der Staten-Generaal, The Hague.

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.

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.