

IZA DP No. 8386

Health Effects of Containing Moral Hazard: Evidence from Disability Insurance Reform

Pilar García-Gómez
Anne C. Gielen

August 2014

Health Effects of Containing Moral Hazard: Evidence from Disability Insurance Reform

Pilar García-Gómez

*Erasmus University Rotterdam,
Netspar and Tinbergen Institute*

Anne C. Gielen

*Erasmus University Rotterdam,
IZA, Netspar and Tinbergen Institute*

Discussion Paper No. 8386
August 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Health Effects of Containing Moral Hazard: Evidence from Disability Insurance Reform*

We exploit an age discontinuity in a Dutch disability insurance (DI) reform to identify the health impact of stricter eligibility criteria and reduced generosity. Our results show substantial adverse effects on life expectancy for women subject to the more stringent criteria. A €1,000 reduction in annual benefits leads to a 2.4 percentage points higher probability of death more than 10 years after the reform. This negative health effect is restricted to women with low pre-disability earnings. We find that the mortality rate of men subject to the stricter rules is reduced by 0.7 percentage points. We hypothesize that the gender difference in health outcomes is due to the reform tightening eligibility particularly with respect to mental health conditions, which are more prevalent among female DI claimants. The evidence for the existence of substantial health effects implies that policy makers considering a DI reform should carefully balance the welfare gains from reduced moral hazard against losses not only from less coverage of income risks but also from deteriorated health.

JEL Classification: I14, H53, I38

Keywords: disability insurance, moral hazard, health, mortality, regression discontinuity

Corresponding author:

Anne C. Gielen
Erasmus School of Economics
Erasmus University Rotterdam
PO Box 1738
3000 DR Rotterdam
The Netherlands
E-mail: Gielen@ese.eur.nl

* Pilar García-Gómez would also like to thank the Netherlands Organisation for Scientific Research for financial support under the Innovational Research Incentives Scheme – VENI. The authors like to thank Pieter van Baal, Dan Black, Hans van Kippersluis, Hale Koc, Erzo Luttmer, Claudine de Meijer, Katherin Mullen, Juan Oliva, Bastian Ravesteijn, Christopher Ruhm, Darjusch Tafreschi, Owen O'Donnell, Jeff Smith, Isabelle Soerjomataram, Jan Stoop, conference participants in the PhD Seminar on Health Economics and Policy, the 2012 Netspar Pension Day, and in the 2014 IZA/SOLE Transatlantic Meeting, and seminar participants in Rotterdam, Lancaster, Universitat Pompeu Fabra, Tilburg and IZA for insightful comments. All errors are our own. The authors acknowledge access to linked data resources by Statistics Netherlands (CBS). Researchers willing to use the data found in this article can apply for access to Statistics Netherlands (CBS). The authors of this paper are willing to offer guidance about the application.

1. Introduction

Many industrialized countries are experiencing an enormous increase in the number of disability insurance (DI) claimants. In the U.S., the share of individuals receiving DI benefits has more than doubled during the last three decades (Duggan and Imberman, 2009; Wise, 2012) and the share of DI spending in total social security expenditures has risen from one in ten dollars in 1988 to nearly one in five dollars in 2009 (Autor, 2011). Since 2009, the ratio of disability beneficiaries to workers in the U.S. has overtaken that of the Netherlands, which was long regarded the ‘world leader’ in DI dependency (Burkhauser and Daly, 2011). Much of the increase in DI roles tends to be attributed to moral hazard and governments have responded by tightening eligibility criteria and/or reducing the generosity of benefits (e.g. Autor and Duggan, 2006; Karlström *et al.*, 2008; Staubli, 2011). Such rationalization in the Netherlands has contributed to the country being knocked off the top of the DI dependency league table.¹

The standard welfare evaluation of DI weighs the gains from insuring income risk against the losses from moral hazard incentives to reduce work effort. Welfare improving reforms encourage the work-able to re-enter employment. Since work capacity is not perfectly observable, a problem which lies at the heart of the moral hazard problem, there is a risk that mistakes are made in any rationalization program or that the process goes too far, such that partially incapacitated individuals for whom work takes an inordinate toll on health are forced back into employment. The health impact may operate, for example, through stress related to re-entry into the labor market, or from work tasks that are physically demanding given certain health conditions. Although there is recent evidence that DI benefit recipients respond to reduced DI benefit generosity and tighter DI benefit eligibility by taking up new employment (Borghans *et al.*, 2014; Moore, 2015), there is no evidence yet about potential health effects *after* such a disability insurance reform.

This paper examines whether adverse health effects may arise as a result of more stringent disability insurance (DI). In particular, we focus on mortality. Other studies investigating

¹ For an overview of the various reforms in the Dutch DI during the 1990s and 2000s, see Van Sonsbeek and Gradus (2013) and Koning and Lindeboom (2015).

health effects of unexpected job loss have documented several health effects such as increased hospitalizations and mortality (Eliason and Storrie, 2009; Sullivan and von Wachter, 2009; Browning and Heinesen, 2012). An unexpected loss of DI benefits might have similar health consequences, as it likely leads to stress associated to financial strain and/or forced re-entry into the labor market, all of which might translate into adverse health effects and possibly an increased mortality risk over time. Evidence for any such effects of DI reform implies a direct loss in welfare. On the other hand, health may also improve if post-reform re-employment improves the cognitive ability of individuals as is often found in the retirement literature (Rohwedder and Willis, 2010; Mazzonna and Peracchi, 2012, 2016; Bonsang *et al.*, 2012).

We provide the first evidence we are aware of on the mortality impact of DI reform by examining the consequences of a Dutch reform in 1993 that entailed medical re-examination of DI recipients, stricter eligibility criteria and benefit cuts.² Identification comes from an age discontinuity – the eligibility criteria were made significantly more stringent for disability recipients aged below 45 on August 1st, 1993. Borghans *et al.* (2014) have shown that many DI beneficiaries experienced benefit cuts following the medical re-examination, and the average reduction was about 10 percent higher for the younger cohort subject to the more stringent criteria. Furthermore, their results also reveal a strong rebound in labor market earnings of 0.62 Euros for each Euro of lost DI benefits, and an increase in other social security benefits of 0.30 Euros for each Euro of lost DI benefits. Hence, the reform clearly had a substantial effect on the labor market position of (former) DI claimants, and as such it may also have had health implications for those affected by the stricter re-examinations.

In this paper we further exploit this discontinuity in the reform by comparing mortality – both in the short and in the longer run – of individuals younger than the 45 year-old cutoff at the time of the reform to the outcomes of individuals older than this threshold. The reform was introduced with little warning and so differences in DI generosity arising from it could not have been anticipated. To estimate the effect on health we use administrative

² Unfortunately, we have no reliable data available on other health outcomes.

panel data providing information on mortality for the total population of Dutch DI claimants.

We find that the reform had a substantial negative impact on the health of female DI claimants. According to the reduced-form estimates, women subject to the stricter DI eligibility criteria had a 1.4 percentage point higher probability of dying within seventeen years after the reform.³ The probability that males die within seventeen years after the reform is estimated to have been reduced by a statistically significant 1.2 percentage points. We find that the negative health effects are concentrated among low earning females. Therefore, we hypothesize that the gender difference in the effect is due to the reform tightening eligibility particularly with respect to mental health conditions, which are more prevalent among female DI claimants, together with a higher job strain related to depression in low skilled occupation.

When allowing for the fact that some individuals were affected more by the reform than others – in terms of benefit cuts – we find that for women a €1,000 reduction in the annual amount of DI benefits led to a 2.4 percentage point increase in the risk of death within seventeen years after the reform. Our results suggest that adverse health effects of tighter DI rules can be empirically important and, in addition to the impact on moral hazard, income insurance and public finances, need to be taken into account in any evaluation of a DI reform.

While the magnitudes of our estimated effects are obviously specific to the particular Dutch reform, the broad conclusions are relevant in a wider context. Health deterioration arising from measures that push (female) DI recipients back to work suggests that the extent of moral hazard may have been overstated. An employment response to DI eligibility rules and generosity is not sufficient to conclude that moral hazard exists. The difficulty in defining and containing moral hazard with respect to DI lies in the fact that disability is not a clear-cut medical condition. Rather it is a threshold on a continuum of work incapacity, and the pain and discomfort associated with work, that is defined by the policymaker

³ Note that this is 14 years after the re-examinations, which took place in 1996/1997 for this relevant cohort.

(Diamond and Sheshinski, 1995). Reforms that tighten the eligibility criteria and so raise the threshold of work incapacity may induce individuals who can only work under excessive duress to take up employment causing health to deteriorate. This argument does not deny the existence of substantial moral hazard in DI, nor does it claim that reforms are unnecessary to reduce moral hazard.⁴ Rather, the point is that setting the threshold of work incapacity involves balancing losses from the Type I errors that represent moral hazard with the losses from Type II errors that arise when individuals who can only work at an excessive cost to their health are denied DI. Given the generosity of the Dutch DI program in the nineties, one would expect the marginal individual affected by the 1993 reform to have been in better health than the marginal DI recipient in a less generous scheme. This suggests that the negative health effects of tightening DI in other countries and periods may be even greater than we estimate for the Dutch setting.

The remainder of the paper is organized as follows. In Section 2, we describe the reform used for identification. Section 3 describes the data, Section 4 describes the empirical strategy, and Section 5 presents the results. Section 6 concludes.

2. Disability Insurance in the Netherlands

2.1 Pre-reform

In the Netherlands, all employees were and are still insured against an earnings loss resulting from an illness or infirmity. In the first year after the onset of a disability individuals are entitled to sickness payments. After this first year, they can apply for DI

⁴ There is a substantial literature on moral hazard in DI. Bound (1989) and Chen and van der Klaauw (2008) found that labor supply of U.S. males over the age of 45 who applied for DI would have been 30 to 40 percent higher if it were not for the availability of DI benefits. Further suggestive evidence for moral hazard is found in the fact that DI applications tend to rise and fall with the unemployment rate (e.g. Bound and Burkhauser, 1999; Autor and Duggan, 2003, 2006). There is also evidence that policies that tighten DI are effective in reducing the number of claimants and bringing people back to work (Autor and Duggan, 2003; Borghans *et al.*, 2014).

benefits.⁵ Note that there is allowance for partial disability and health insurance is not tied to the receipt of DI benefits.

The amount of DI benefit depends on the "degree of disability", which equals the percentage difference between earnings before disability and the potential earnings of the applicant, which prior to the 1993 reform was determined in three steps. First, the applicant was examined by a medical doctor, who compiled a list of job demands that the applicant could still meet.⁶ Second, this list was compared with a dictionary of occupations that specified the job demands, as well as the required education level, for each occupation in order to identify a set of occupations the applicant would be able to perform. Only occupations that were no more than two education levels (on a seven point scale) below the level demanded in the applicant's previous occupation were considered. Finally, the applicant's potential earnings capacity was defined as the mean wage of the 5 highest paying occupations in the set of feasible ones, with the further proviso that at least 10 workers should be currently engaged in each of the occupations in the applicant's region (though there should not necessarily be that number of vacancies).⁷ When 5 suitable occupations with at least 10 workers could not be identified, the potential earnings capacity was set at 0. The degree of disability was then defined by the percentage loss in earnings due to disability, i.e. the difference between prior labor earnings⁸ and the potential earnings capacity divided by prior labor earnings, and grouped into 8 categories varying from 0-15% to 80-100%. The replacement rate was then determined by these categories (see Table 1).

⁵ Further institutional background information on the Dutch disability act and recent trends and patterns in Disability Insurance enrollment in the Netherlands can be found in Bovenberg (2000), García-Gómez *et al.* (2011) and Koning and Lindeboom (2015), respectively.

⁶ The complete list includes 27 physical job demands (such as "lifting," "kneeling," and "ability to deal with temperature fluctuations") and 10 psychological skills (such as "ability to work under time pressure," "ability to perform monotonous work," and "ability to deal with conflict").

⁷ For this procedure, the Netherlands was divided in 5 regions and in 16 "start regions." Alternative jobs in the "start regions" had to be considered first. Only if less than 5 suitable occupations were available, alternative jobs in the neighboring regions (within one of the main 5 regions) could be considered.

⁸ Prior labor earnings were subject to a cap, which was about €36,000/year in 1999.

2.2 *The reform*

The 1993 reform was designed to lower the generosity of disability benefits and to reduce the number of (full) disability claimants.⁹ With this aim, the criteria to define the earnings capacity were changed in two respects.¹⁰ First, eligibility was only based on objective medical information, not the doctor's judgment. This required that a functional work limitation was clearly observable, and could be linked directly to a medical diagnosis. As a result, disabilities related to mental health problems became more difficult to prove. Second, the criteria for identifying suitable occupations were relaxed, such that: (i) all education levels could be considered, (ii) only 3 occupations (rather than 5) with at least 10 workers employed were required to calculate the mean potential earnings, and (iii) the geographic region within which there should be employment in the occupations was expanded roughly threefold.¹¹ These changes increased the probability of finding a higher-paid occupation that the applicant could still perform, and thereby raised the potential earnings capacity. Furthermore, it became less likely that the potential earnings capacity was set at 0 because insufficient occupations could be identified. For any individual who was enrolled in DI, the potential earnings capacity under the new criteria was always higher than under the old criteria and, hence, DI benefits were lower.

The new criteria for determining the potential earnings capacity applied both to new applicants and to existing claimants aged younger than 50 at the time of the reform (August 1st, 1993). Existing claimants were to be re-examined, which was scheduled to take place by age cohort from 1994 onwards: those aged <35 on the 1st of August 1993 in 1994, 35-40 in 1995, 41-44 in 1996/1997, and 45-49 in 1997-2001. However, shortly before the re-examinations of the 45-49 cohort started, political pressure¹² led to a parliamentary motion

⁹ The formal name of the 1993 DI reform is "Terugdringing Beroep op Arbeidsongeschiktheidsverzekeringen (TBA)," which roughly translates as "Reducing claims on disability insurance".

¹⁰ Another important change of the 1993 DI reform was the introduction of an age- and duration-dependent benefit for new DI applicants. However, these changes did not apply to those already receiving disability benefits as of August 1993, i.e. the group that we study here.

¹¹ Now, rather than occupations in only one of the 16 "start regions", all available jobs within the main region where the individual was residing (out of 5 main regions) could be considered.

¹² This had to do with the weak labor market position of older individuals and the difficulties that those aged 45-49 would experience in finding a job if their DI benefit would be reduced or ended. Relaxing the eligibility criteria for this group would prevent that these individuals would end up in poverty.

passed in November 1996 ruling that this cohort would be assessed on the pre-reform, more lax criteria. This created a discontinuity in DI stringency between the cohort under the age of 45 and those aged 45-49 that we exploit to estimate the effects of reduced generosity on health outcomes.¹³

3. Data

3.1 Data sources

We extract our data from several large administrative databases maintained by Statistics Netherlands. Individuals can be matched across the various data files by a Random Identification Number (RIN), which is an encrypted Dutch equivalent of the U.S. Social Security number.¹⁴

We use administrative data on all disability benefits recipients aged 15-64 in the Netherlands for the period 1995-2005. These data originate from the organizations that administered the disability benefits at the time. They include information on entry and exit dates of a disability spell, the degree of disability (in categories), labor earnings prior to DI entry and disability benefit payments. Unfortunately, there is no reliable or consistent information about the medical condition that led to enrollment.

Demographic characteristics of the claimants are obtained from the municipal registries. These contain information on date of birth (year and month), nationality, marital status and place of residence for all residents of the Netherlands.

Information on other sources of income are obtained by merging five different administrative datasets: earnings from paid employment of all employees, self-employment earnings, unemployment benefits, general assistance, and receipt of other

¹³ We cannot exploit the discontinuity at age 50 since the re-examinations of those ages 48 and 49 were gradually phased out because of the long backlogs that arose at the disability office.

¹⁴ These data can be accessed via a remote-access computer after a confidentiality statement has been signed.

types of social assistance (from about 30 relatively minor programs). Information about income from paid employment and self-employment comes from the tax authorities and social insurance records, whereas the data on social assistance are provided by the organizations that administer these programs. All these files are available from 1999 onwards.

Last, we obtain information on death from mortality registries covering the entire Dutch population. Mortality records are available for the period 1996-2010 from the reports filed by a medical examiner or pathologist.

3.2 Sample definition

Our baseline sample consists of all individuals who (i) received DI benefits on August 1st, 1993, (ii) were aged between 40 and 50 on that date, (iii) were still on DI as of January 1st, 1996, and (iv) are native Dutch. The first restriction follows from the fact that the discontinuity in DI generosity only applies to disability claimants enrolled at the time the reform went into effect. The second restriction is made to create a sample of individuals who are relatively comparable in all other respects except for the assessment of earnings capacity made during re-examination and the consequent difference in DI generosity. The standard criterion by Imbens and Kalyanaraman (2012) for setting the window of a regression discontinuity (RD)¹⁵ would suggest taking an estimation sample only including claimants aged between 43 and 47. However, with this age window there is insufficient power: at most 0.30 for males and 0.25 for females. Therefore we have extended the age window to claimants aged 40 to 50. Results for the smaller sample are rather comparable to those for the wider sample (see section 5.2 for more details). The third restriction is imposed by the fact that data on DI status are available only from 1996.¹⁶ From those observed on DI in January 1996, we select those whose records show they have been claiming at least since 1st January 1993. Given that the re-examinations of the individuals in our sample did not start before 1996 and the parliamentary motion excluding the cohort

¹⁵ More information about our empirical strategy follows in section 4.

¹⁶ For 1995 we only have information on those who left DI, not on those who stayed on DI. Hence, information on continuing DI spells is only available from January 1st, 1996.

aged 45 and above from the more stringent criteria was only passed in November 1996, we do not expect any differential attrition to have taken place around the age of 45 prior to January 1st, 1996.¹⁷ This assumption is supported by Appendix Figure A1 where we show that there is no differential exit out of DI for individuals aged below and above 45 in 1995.

Finally, the last restriction comes from the fact that we have to drop all non-native Dutch from our sample because of two major problems with certain groups of immigrants that potentially invalidate the identification strategy. First, more than a proportional share of Turkish immigrants is registered as being born in January, which reflects inaccurate birth registration and is problematic for our identification based on date of birth.¹⁸ Second, there are large differences in the share of immigrants from the East Indies across cohorts on either side of the threshold determining exposure to the more stringent test.¹⁹ We cannot rule out that there are similar problems with other groups of immigrants. We decide to take a conservative approach and drop all non-native individuals from our sample given that there are differences in health and life-expectancy between individuals from different ethnic backgrounds. However, we will show in section 5.2 that the results are not changed when non-native Dutch are included in the sample. After these sample restrictions, our baseline sample in 1996 contains 91,089 males and 46,670 females in 1996.²⁰

¹⁷ Please note that throughout this paper we refer to cohorts by their age at the time the reform went into effect (i.e., as of August 1st, 1993).

¹⁸ It was common for families in rural areas of Turkey not to register new-born children immediately after they were born, but rather to wait until more children were born and then register them all at once. Very often, parents did not remember the exact birth date of all their children while at the registration office, and registered them as being born on January 1st.

¹⁹ Directly after the Second World War (WWII) a large inflow of immigrants from the East Indies who fought in the Royal Netherlands East Indies Army came to the Netherlands and had children shortly after arrival. These people are aged 47 at the time of the reform, which creates important differences in the composition of the group aged 40-44 compared to the group aged 45-49.

²⁰ About 3 percent of the individuals in our sample appear in more than one disability record in a given month. These observations are dropped because it is not clear whether they are administrative/coding errors or whether they arise from individuals who are entitled to multiple DI benefits because they were employed in multiple jobs before the onset of disability. Since there is no discontinuity at age 45 in the probability of someone having multiple disability records, we believe that omitting/including those with multiple records would not substantively change our results.

3.3 Summary statistics

Table 2 presents a summary of our variables of interest. Panel A shows the characteristics of the individuals in our sample as of 1st January 1996 before they were re-examined. On average, the claimants in our sample have spent 7-8 years on disability insurance at the time the reform was implemented. The majority of the sample is fully disabled (between 80 and 100% earnings reduction). This holds in particular for women.

Panel B presents the mortality rate for the last year of our observation period. The mortality rate was higher for men. By 2010, 15.6 percent of the male, but only 11.2 percent of the female, DI claimants had died.

4. Empirical Strategy

The 1993 DI reform in the Netherlands has had a significant impact on the average benefits that a DI claimant received. On average, disability benefits were reduced by €1,549 and €607 for men and women under the age of 45, respectively (see Figure 1, Panel I). This amounts to a drop of about 13 percent and 7 percent of total disability insurance benefits for men and women, respectively. These first stage effects are in line with those in Borghans *et al.* (2014). Note that these are average reductions in DI benefits; some individuals experienced a decrease in benefits, whereas others experienced no change in benefits at all (or even an increase). About 12-13 percent of men and women under the age of 45 faced a reduction in DI benefits after the re-examinations, contrary to only 3.4 percent and 2.5 percent of men and women aged 45 and older, respectively (see Table 2). The more stringent re-examinations not only reduced the generosity of the DI program for individuals remaining on DI, it also led to additional exit out of disability of about 5.6 and 6.9 percentage points for men and women, respectively (see Figure 1, Panel II). In fact, we observe that DI exits account for more than 90 percent of the cases where a DI claimant experienced a drop in DI benefits. Hence, exit from the program is an important effect of the DI reform, and the differential impact on changes in DI benefits is likely due to the fact that pre-reform benefits for men exceed those of women (as men's pre-disability earnings

were on average 45% higher than those of women, see Table 2). We exploit these first stage effects to examine whether the reform had any impact on health outcomes. Note that we are identifying the effect of a change in *disability benefit* income on health outcomes; it is not the effect of a change in *total* income on health as Borghans *et al.* (2014) have shown that total income was only slightly reduced for men and was unchanged for women since the reduced benefits were to a large extent compensated by increased employment earnings or additional income from other benefits.

In our regression models, treatment is indicated by a dummy equal to 1 for those subject to the stricter re-examinations (i.e. those aged under 45). Our set of controls include age in months as of the 1st of August 1993, an interaction between the treatment indicator and age measured in months younger/older than 45 in August 1993 to allow the age trend to differ on either side of the discontinuity, 6 dummies for the degree of disability on January 1st, 1996, pre-DI earnings, 39 regional dummies, marital status and duration on DI at the start of the reform. To control for seasonal patterns in mortality by month of birth (Doblhammer and Vaupel, 2001), we also control for 12 month of birth dummies. Finally, we include a dummy for those born between May 1945 and January 1946, and a dummy for those born between February 1946 and January 1947 to allow for potential differences in the composition of the baby boom generation and for the possibility that those born shortly after World War II differ in health status from other cohorts because of conception during the last months of the war that were particularly harsh in the Netherlands (Scholte *et al.*, 2012).

One key identifying assumption in this study is that the stringency of the criteria applied during the DI re-examinations is the only discontinuous change relevant to mortality at the cutoff age of 45. Although the data do not allow us to test this, we are not aware of any other policy change that would create a discontinuity around this age. To give more credibility to our results, we test whether there are any discontinuities around age 45 in our predetermined control variables by regressing all covariates used in the analysis on the treatment indicator in a reduced-form RD specification. Under the null hypothesis that there is no effect, from 51 placebo regressions we would expect to find a significant

coefficient for only 3 (at 5%) or 5 (at 10%) of the covariates. We find a significant effect for only 3 of them.²¹

Another key assumption guaranteeing the validity of our approach is the fact that individuals should not be able to influence the running variable, i.e. age at August 1st, 1993. Since individuals cannot easily change their actual or reported age in the Netherlands, we do not expect any manipulation of reported age. This is further confirmed by Appendix Figure A2, which shows a histogram of the DI claimant's age at the time of the reform. Although the Figure displays some variation, there is no substantial change in claimant's age around the age cutoff. Also a McCrary (2008) density test provides further evidence for a smooth density around age 45 (p-value=.114). Note that we do observe a drop in the histogram around age 47.5 which is due to the famine resulting from the hunger winter which took place in the Netherlands during World War II. Hence, this drop is not related to the DI reform. Nevertheless, we control for this event in our regressions (as described above).

5. Results

5.1 Impact on mortality

Figure 2 shows the probability of death by 2010 for the individuals in our sample. There is an upward jump in the male mortality rate at the age of 45, which corresponds to the age from which the more generous DI reassessment applied, but a downward jump in female mortality. Men aged below 45 who were subject to the stricter regime had a 1.2 percentage point lower probability of having died by 2010 than older men, controlling for age. Younger women that experienced the tighter reassessment had a 1.4 percentage point higher probability of death, which corresponds to a relative effect of about 13 percent. Estimating a reduced-form RD model for mortality in earlier years (Figure 3) we find the positive effect on mortality for women and the negative effect for men to be consistent over time, where the magnitude of both increases over time and reaches significance in the last

²¹ The full results are available in Table A1 in the Appendix.

two years (see also Appendix Table A2 for the complete estimation results). Interpreting these estimates as causal effects would suggest that tightening the eligibility criteria and reducing benefits raised the mortality rate of women, but reduced that of men. In Section 5.2 we show that these results are robust to various robustness checks. Note that Borghans *et al.* (2014) have shown that the reform induced women to take up new employment and also to apply for other benefits at a higher rate than men. This suggests that women may have been affected more by the reform than men, and this might explain the gender difference in mortality. In addition, the gender difference in mortality may be the result of different social norms to working or to different health problems for men and women. We will discuss the potential explanations for the gender difference in health outcomes in detail in section 5.4.

The reduced-form RD point estimates shown in Figure 3 and Table A2 show the average effect of the reform on mortality. However, some individuals were affected more by the reform than others. That is, some individuals lost all their DI benefits because they were considered ineligible under the new criteria, whereas others experienced only a marginal reduction in their benefits. We exploit this variation in DI benefit changes in an IV standard fuzzy RD specification, where we scale our reduced-form estimates by the amount of the reduction in disability benefits.²² The IV estimates in Table 3 show that the mortality effects are larger for those who experienced a larger drop in DI benefits. For example, a €1,000 reduction in DI benefits led to an increase in the cumulative mortality rate by 2010 among women of 2.4 percentage points, which is a relative increase of 23 percent.²³ For men we find that a €1,000 reduction in DI benefits has led to a 0.7 percentage point (i.e. 6 percent) reduction in mortality by 2010.

It is important to note that although our mortality effects only appear from 2009 onwards, they are not caused by any of the later DI reforms. As Koning and Lindeboom (2015) carefully describe, after 1993 there have been several DI reforms which aimed at (i)

²² This is the change in DI benefits due to the reform. Note that benefits are set to missing for those already deceased by January 1999.

²³ Unfortunately, we have insufficient observations for each cause of death to investigate whether this result is driven by certain specific causes of death.

enhancing employer incentives to avoid an inflow of employees in disability insurance, (ii) increasing screening for disability and (iii) tightening eligibility for continued DI benefit receipt. Although these reforms could have affected people who continued to be on DI after the re-examinations, none of these reforms had a *differential* impact on the 40-44 year old cohort versus the 45-49 year old cohort in our study.²⁴ In fact, there was no particular discontinuity in treatment for any age in our 40-50 cohort. As such, these later DI reforms are no threat for our identifying assumptions. Furthermore, one might wonder whether the long-term mortality effects could be due to recent changes in early retirement schemes, since our sample individuals who were aged 45 in 1993 are around 61/62 in 2009/2010, so well before the official retirement age of 65. However, although there has been substantial changes in early retirement schemes since the mid-1990s and 2000s (see for example Euwals *et al.* (2011) for more details), none of these changes affected our 40-44 year cohort differently than it affected our 45-49 year cohort.²⁵

5.2 Robustness of the results

In Table 4 we present the results of various sensitivity analyses. Panel A shows the results of reduced-form RD estimations without any control variables. Clearly, the results are not sensitive to inclusion of the controls. The mortality effects remain -0.012 for males and 0.014 for females, just as was found when controls were included in the regression (Figure 3). In Panel B we include a quadratic in age in the model. The results remain mostly unchanged. Panel C shows the outcomes when all immigrants are included in the sample. The results for women remain unchanged. For men, the significant mortality effects disappear. This is consistent with differences in the share of immigrants (at least from East Indies) between age groups, and lower life expectancy of non-natives compared to natives. In addition, measurement error in the month and year of birth would bias our estimates towards zero.

²⁴ See Koning and Lindeboom (2015) for more details.

²⁵ Ages refer to those on August 1st, 1993. By 2009, these ages would be 56-60 and 61-65.

We also perform several placebo tests. In Panel D we re-estimate our models on a sample of disability claimants aged 50-59, applying an artificial age discontinuity at age 55. Note that DI claimants aged 50 and above were not re-examined at all, nor did any of the new eligibility criteria apply to them. Hence, we should not find any significant effect for these placebo estimations. The results in panel D confirm this. A final placebo test concerns the estimation of our model on a sample of individuals who were not receiving DI at the time of the reform. The results in panel E show that there is no discontinuity in mortality for this sample, which provides further confidence on the absence of any other confounding factors that might occur at the cutoff age 45. This confirms the validity of our approach.

As mentioned in section 4, we choose a larger bandwidth than would be optimal according to the Imbens-Kalyanaraman criterium given the lack of power for the smaller bandwidth. One may be concerned that the health effects are therefore driven by the inclusion of individuals further away from the threshold so they are the result of age differences. For the results for women we argue that it is unlikely to be the case as treated individuals are younger and this would imply lower mortality for the treated group. However, for men we cannot rule out that this is the case. However, in Figure 4 we investigate how sensitive our results are to the chosen bandwidth size. The figure illustrates that the point estimates of our models are quite robust to bandwidth size.²⁶ More specifically, the results remain rather stable after the optimal bandwidth of 24 months as defined by the Imbens-Kalyamaraman procedure.²⁷ The standard errors of the estimates, however, are sensitive to the bandwidth, and the figures show that as the power increases (larger bandwidth includes larger number of observations), the estimates also become more precise.

5.3 Are the effects driven by changes in employment or total income?

Our results show that changes in the DI generosity have important health effects on DI recipients. The negative health effects among women are likely to be driven by increased stress related to new employment situation or reduced benefit income, and/or occupational

²⁶ Only for very small bandwidths they are quite different, but one should be careful interpreting these point estimates since also the standard errors are very large.

²⁷ Although they are slightly smaller for female mortality between bandwidth of 24-36 months.

diseases if the person enters into paid employment. The reduced generosity of this reform has been found to have a large effect on employment and only a very small effect on total income. Borghans *et al.* (2014) have shown that participation in employment went up by 2.9 percentage points after the reform and that also total income changed, albeit only marginally: on average, for every euro lost on DI benefits, about 92 cents is compensated through increased labor earnings or other benefit income. We exploit our information on employment status and total income to disentangle the importance of both in explaining the mortality effects.

We re-estimate our IV models using two additional scaling variables: a dummy for whether the individual is in paid employment in 1999 and individual total income in 1999 (i.e. after the re-examinations were finished). Panel A of Table 5 shows that conclusions are similar if we use employment status as the scaling variable and the treatment dummy as the instrument. There are health costs (gains) for those women (men) that get into paid employment due to the stricter eligibility criteria. On the other hand, income changes driven by the reform do not seem to explain changes in mortality (Panel B). For women, the reform did not have an effect on total income (no first stage effect). This is in line with the finding by Borghans and co-authors (2014) that individuals are able to fully compensate a drop in DI benefits via higher labor market earnings and/or other benefits. For men, the reform had a small effect on total income, but a €1,000 reduction in total income is not associated with a statistically significant mortality effect. Therefore, the previous mortality effects for men and women must be driven by individuals who left DI and moved into employment due to the more stringent DI criteria, and not by those who stayed on DI with reduced benefits.

5.4 Why did the reform only harm the health of women?

Our results have shown clear gender differences in the health effects of the Dutch DI reform. It is possible that the reform affected men and women in a different way. We have already seen (see Figure 1) that the average reduction in DI benefits due to the reform is larger for men than for women (1,549 Euros versus 607 Euros), but that the probability to leave DI after the re-examinations is larger for women (6.9 percentage points versus 5.6

percentage points). Given that we observe that terminations of DI eligibility (i.e. DI exits) account for more than 90 percent of the cases where a DI claimant experienced a drop in DI benefits, this gender difference in exits strongly suggests that women are more affected by the reform than men. Further evidence is provided by Borghans *et al.* (2014), who have shown that both the probability to take up employment as well as the probability to obtain other benefits increased significantly more for women than for men.

One of the reason for the gender-specific impact of the reform could be related to gender-specific health problems. One of the implications of the new criteria is that disabilities related to mental health problems became more difficult to prove than physical health problems (see section 2). Women are more likely to suffer from mental problems (37% of females vs. 28% of males), while men are more likely to suffer from physical health problems, like musculoskeletal (31% of males and 26% of females) and cardiovascular (8% of males vs. 3% of females) problems (UWV, 2006).²⁸ Hence, these differences in health problems might lead to differential exit from DI across both genders: one would expect the average man who left DI to be healthier than the average woman who left DI, as the share of reclassifications due to a health recovery in total male exits is larger than for women (since for women the share of exits due to ineligible type of illness – i.e. mental health problems that are difficult to diagnose – is larger). This could explain the gender differences as employment may be more hazardous when initial health is poor. Similarly, one might also expect heterogeneous effects by type of health condition. Unfortunately, information about type of illness is missing in our data so we cannot test for this.

In order to ascertain what can explain the gender differences, we pool observations for men and women, and make a distinction by three other variables that may drive these results: degree of disability (full/partial), disability duration and earnings level. The results are shown in Table 6.²⁹

²⁸ Numbers are averages calculated for period after the reform (1994-2005).

²⁹ In Appendix Table A3 we investigate the characteristics of “compliers”, i.e. those individuals who are affected on the margin by the stricter re-examination. Hence, compliers refer to those individuals who experience a decrease in their degree of disability following the stricter re-examination criteria but would not

First, women are more likely to be *fully* disabled compared to men (see Table 2). One would expect that fully disabled would suffer more from a DI reform that decreases the amount of benefits and encourages disabled individuals to go back to work than partially disabled, as the labor market opportunities for the former are expected to be worse. The DI reform raises the probability of mortality by 2010 for both fully and partial disabled women, while it decreases the probability of mortality by 2010 only for fully disabled men (Panel A of Table 6). In addition, the estimated effect is statistically different for males and females in the two subgroups (partial and fully disabled). This suggests that the gender-specific findings cannot (only) be attributed to differences in the level of disability between genders.

Second, we investigate the time spent on DI at the time the reform came into effect. On the one hand, individuals who have been on DI for longer may suffer more from forced labor market re-entry as they have been detached from the labor market for longer. On the other hand, the health of long-term disabled may have improved while on DI, therefore they may well be able to find re-employment without this having adverse health effects. From Table 2 we know that women have slightly shorter DI spells than men. Panel B of Table 6 shows separate results by time spent in DI. We find that the mortality effects for men are concentrated among long-term disabled, whereas there are negative effects for women who are either long-term or short-term disabled. Moreover, the difference in the direction of the effect between males and females remains among the long-term disabled. This suggests that gender-specific findings can also not (only) be driven by differences in the time spent on DI across genders.

There are large differences in the labor market history and opportunities between men and women, as women tend to be low wage earners. Therefore, our results could be well driven

have experienced a decrease in their degree of disability under the old and more lenient criteria. Since our results suggest that partially disabled women are more likely to experience a reduction in their benefits than the female sample as whole, in Table 6 we investigate further whether the gender-specific results are related to initial degree of disability.

by lower wage earners (mainly women) having to return to more health hazardous jobs compared to higher earners (mainly men). In Panel C of Table 6 we separately estimate the effect of the reform for those below and above the median level³⁰ of labor market earnings (measured before entry in DI). We find that the negative health effects are indeed concentrated among low earning women, whereas the positive health effects for men are concentrated among high earning men. Figure 5 shows the reduced-form RD effects on mortality over time for the four different groups (low earnings females, high earnings females, low earnings males, and high earnings males). We find that the positive health effects on high earning males affected by the reform are apparent in the short-run (from 1998). Moreover, the estimated effect on high earning females is similar both in size and sign, and not statistically different to the estimated effect on high earning males, although it is never statistically significant. The lack of statistical significance can be due to the low number of females that are among the high earners, as only 24% of females have pre-DI earnings above the median. The differences between high earning males and low earning females are clear since the early 2000s. This suggests that part of the gender differences is due to differences in the skill level between males and females in our sample. However, this cannot be the only explanation as the reform did not have any effect on low earning males. In addition, these results pose some equity concerns on the effects of the DI reform, as the health gains are concentrated among the high-earners while the losses are concentrated among the low-earners.

Furthermore, gender differences in occupational sorting may be responsible for these differential health effects. LaMontagne *et al.* (2008) have shown that job strain related to depression is mostly concentrated in low skilled occupations, and national statistics show that female recipients in the relevant age group are about 10 percentage points more likely to suffer from mental health problems compared to males (UWV, 2006). Hence, this suggests that the combination of the type of health problems and the occupational position makes women more vulnerable to the reform. Unfortunately, type of illness is not registered in our data and we cannot test this hypothesis directly.

³⁰ Note that we use the overall median wage, not gender specific. When using gender specific medians, we could not separate the gender effects from the earnings effect.

Finally, the differential gender effect on mortality may be the result of different social norms to working. We know from earlier studies that leaving employment after job loss has severe and long-lasting effects on mortality (Eliason and Storrie, 2009; Sullivan and Von Wachter, 2009; Bloemen et al., 2015), and that these effects are larger the stronger is the social norm to working (Stutzer and Lalive, 2004). If the effects of losing a job are symmetric, but opposite, to the effects of regaining employment, then one may argue that forced exit from disability and the subsequent re-employment may have beneficial effects on mortality. These effects are expected to be largest for men, as the social norm to working is strongest for them. For women, the positive mortality effects from re-employment may be smaller as the social norm to working is lower for women, and hence the positive effects may be outweighed by the negative effects following from the type of health problem and occupational position of women as was discussed above. Unfortunately, the data do not allow us to test the role of social norms in explaining the different health outcomes across men and women.

6. Conclusion

In this paper, we investigate whether disability insurance reform aiming to contain moral hazard may have additional welfare effects via changes in people's health. The standard assessment of welfare effects of DI balances the gains from insuring incomes against the losses from moral hazard. Welfare improving reforms aim to encourage the work-able to become re-employed. However, when this is pushed too far and even less work-able are forced back to work this may come at a long term cost to their health. Such adverse health effects are clearly undesirable from a human perspective. Although there is ample evidence that DI benefit recipients respond to reduced DI benefit generosity and tighter DI benefit eligibility by taking up new employment, there is no evidence yet about potential health effects *after* such a disability insurance reform. A better understanding of such effects is crucial for designing effective public policy

We exploit a cohort discontinuity in the stringency of the 1993 Dutch disability reforms to obtain causal estimates of the effects of decreased generosity of disability insurance (DI) on health outcomes of existing DI recipients. We find that the reform in Dutch DI has had important health effects which last for more than 15 years after the reform. Women aged 44 and younger which are subjected to the stricter re-examinations experience a worsening in their health status over time as their mortality was increased by 1.4 percentage points by 2010. For men, on the other hand, mortality rates were reduced by 1.2 percentage points. Therefore, there were clear gender differences in the effect of the reform, and these gender differences are even more pronounced in the IV results. A plausible explanation is that the reform affected men and women in a different way, as mental health problems, which were more prevalent among women, became more difficult to prove than physical health problems. This would imply that women leaving DI were on average in worse health than men, and therefore more sensitive to the occupational risks. Alternatively, different social norms to work apply for men and women, which may explain the gender different in health outcomes. Unfortunately, our data do not allow us to test these explanations, and this is left for future research.

We investigate if other factors, which are correlated with gender, may be responsible for the gender differences. Our results show that the gender-specific findings cannot (only) be attributed to differences in the level of disability neither to differences in the time in DI between genders. However, pre-DI earnings differences play an important role. We find that the health gains are concentrated among high earning males, while the health costs are mostly borne by low earning females. Job strain related to depression seems to be concentrated in low skilled occupations (LaMontagne, 2008). This suggests that the combination of the type of health problems and the occupational position makes women more vulnerable to the reform.

Regardless of the specific driving mechanisms behind these gender differences, our results illustrate that there were clear winners and losers from this reform. We believe this is a relevant lesson applicable to other institutional settings and DI reforms. On the one hand, tightening of the eligibility criteria to reduce moral hazard cannot only increase labor force participation but also improve population health if targeted to the appropriate groups. On

the other hand, if the criteria do not allow benefits entitlement to individuals in some vulnerable groups, the incentives to go back to work may cause irreversible damage to their health. In order to advance in the identification of the groups that can benefit from a DI reform and the groups that society needs to protect, future research should focus on better identification of the effects of DI policies on individuals with different health conditions and in different occupational groups.

All in all, our results indicate that the moral hazard debate should change dramatically: the discussion should not only focus on the tradeoff between providing disability insurance and making people go back to work, but also on other important behavioral effects that may not only compensate the savings from reduction in the DI generosity, but entail additional costs in other government budgets and the society as a whole.

References

- Autor, David H. 2011. The unsustainable rise of the disability rolls in the United States: Causes, consequences, and policy options. Working Paper no. 17697, National Bureau of Economic Research, Cambridge, MA.
- Autor, David H., and Mark G. Duggan. 2003. The rise in the disability rolls and the decline in unemployment. *Quarterly Journal of Economics* 118(1): 157 - 206.
- Autor, David H., and Mark G. Duggan. 2006. The growth in the social security disability rolls: A fiscal crisis unfolding. *Journal of Economic Perspectives*, 20(3, Summer), 71-96.
- Bloemen, Hans, Stefan Hochguertel, and Jochem Zweerink. 2015. Job loss, firm-level heterogeneity and mortality: Evidence from administrative data, *IZA Discussion Paper Series*, no. 9483, IZA, Bonn.
- Bonsang, Eric, Stéphane Adam, and Sergio Perelman. 2012. Does retirement affect cognitive functioning? *Journal of Health Economics*, 31, 490-501.
- Borghans, Lex, Anne C. Gielen, and Erzo F.P. Luttmer. 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, John. 1989. The health and earnings of rejected disability insurance applicants. *American Economic Review* 79(3): 482-503.
- Bound, John, and Richard V. Burkhauser. 1999. Economic analysis of transfer programs targeted on people with disabilities. In *Handbook of Labor Economics* Volume 3C, ed. Orley Ashenfelter and David Card. Elsevier.
- Bovenberg, A. Lans. 2000. Reforming social insurance in the Netherlands. *International Tax and Public Finance* 7(3): 345-368.
- Browning, Martin, and Eskil Heinesen. 2012. Effect of job loss due to plant closure on mortality and hospitalization. *Journal of Health Economics*. 31: 599-616.
- Burkhauser, Richard V., and M.C. Daly. 2011. *The declining work and welfare of people with disabilities: What went wrong and a strategy for change*. Washington DC: American Enterprise Institute Press.
- Chen, Susan, and Wilbert van der Klaauw. 2008. The work disincentive effects of disability insurance program in the 1990s. *Journal of Econometrics*, 142(2): 757-784.

- Diamond, Peter, and Eytan Sheshinski. 1995. Economic aspects of optimal disability benefits. *Journal of Public Economics* 57(1): 1-23.
- Doblhammer, Gabriele, and James W. Vaupel. 2001. Lifespan depends on month of birth. *Proceedings of the National Academy of Sciences of the United States of America*, 98(5): 2934-2939.
- Duggan, Mark G., and Scott A. Imberman. 2009. Why are the disability rolls skyrocketing? The contribution of population characteristics, economic conditions, and program generosity. In *Health at Older Ages: The Causes and Consequences of Declining Disability*, D.M. Cutler and D.A. Wise (eds.), Chicago: University of Chicago Press, 337– 379.
- Eliason, Marcus, and Donald Storrie. 2009. Does job loss shorten life? *Journal of Human Resources*, 44(2): 277-302.
- Euwals, Rob, Annemiek van Vuren, and Daniel van Vuuren. 2011. The decline of early retirement pathways in the Netherlands. *CPB Discussion Paper*, no. 179, CPB, The Hague.
- García-Gómez, P., H.-M. von Gaudecker, and M. Lindeboom. 2011. Health, disability and work: Patterns for the working age population. *International Tax And Public Finance* 18(2): 146-165.
- Imbens, Guido W., and Karthik Kalyanaraman. 2012. Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies* 79(3): 933-959.
- Karlström, Anders, Mårten Palme, and Ingemar Svensson. 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, Pierre W.C. and Maarten Lindeboom (2015). The rise and fall of disability enrollment in the Netherlands. *Journal of Economic Perspectives*, 29 (2), 151-172.
- LaMontagne, Anthony D., Tessa Keegel, Deborah Vallance, Aleck Ostry, and Rory Wolfe. 2008. Job strain – Attributable depression in a sample of working Australians: Assessing the contribution to health inequalities. *BMC Public Health*, 8: 181-189.
- Mazzonna, Fabrizio, and Franco Peracchi. 2012. Ageing, cognitive abilities and retirement. *European Economic Review*, 56(4), 691-710.
- Mazzonna, Fabrizio, and Franco Peracchi. 2016. Unhealthy retirement. *Journal of Human Resources*, forthcoming.

- McCrary, Justin. 2008. Manipulation of the running variable in the regression discontinuity design: A density test, *Journal of Econometrics* 142(2): 698-714.
- Moore, Timothy J. 2015. The employment effects of terminating disability benefits. *Journal of Public Economics*, 124: 30-43.
- Rohwedder, Susann, and Robert J. Willis. 2010. Mental retirement. *Journal of Economic Perspectives* 24(1): 119-138.
- Scholte, Robert S., Gerard J. van den Berg, and Maarten Lindeboom. 2012. Long-run effects of gestation during the Dutch winter famine on labor market and hospitalization outcomes. Working Paper no. 6307, Institute for the Study of Labor (IZA), Bonn, Germany.
- Staubli, Stefan. 2011. The impact of stricter criteria for disability insurance on labor force participation. *Journal of Public Economics* 95 (9-10): 1223–1235.
- Stutzer, Alois, and Rafael Lalive. 2004. The role of social work norms in job searching and subjective well-being. *Journal of the European Economic Association* 2(4): 696-719.
- Sullivan, Daniel, and Till von Wachter. 2009. Job displacement and mortality: An analysis using administrative data. *Quarterly Journal of Economics* 124(3): 1265-1306.
- UWV. 2006. *Kroniek van de sociale verzekeringen 2006*. Amsterdam.
- Van Sonsbeek, Jan-Maarten and Raymond H.J.M. Gradus. 2013. Estimating the effects of recent disability reforms in the Netherlands. *Oxford Economic Papers* 65(4): 832-855.
- Wise, David A. 2012. *Social security programs and retirement around the world: Historical trends in mortality and health, employment, and disability insurance participation and reforms*. Chicago: The University of Chicago Press.

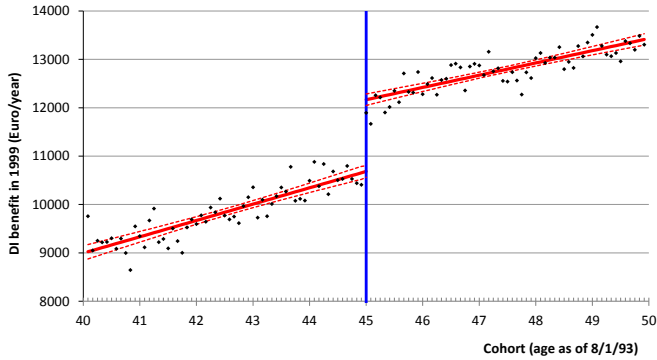
Figures and Tables

Figure 1: First stage effect of DI reform on DI outcomes

I. Effect of DI reform on annual DI benefit amounts received (in 1999, in Euros)

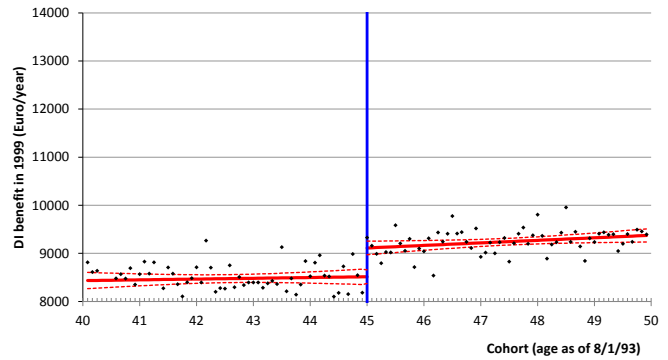
A. Males

Estimate of the discontinuity:
-1549.04 (91.92)



B. Females

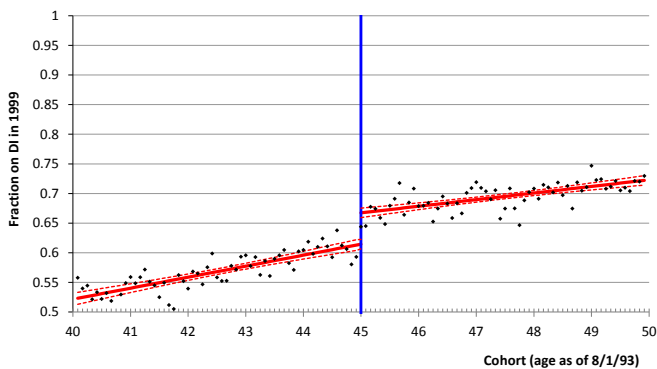
Estimate of the discontinuity:
-606.625 (110.00)



II. Effect of DI reform on participation in DI in 1999

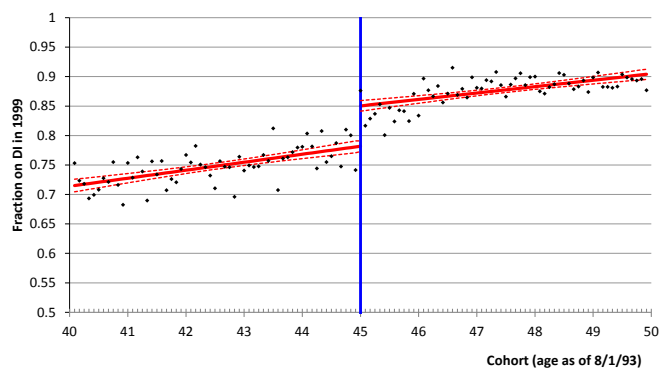
A. Males

Estimate of the discontinuity:
-0.056 (0.006)



B. Females

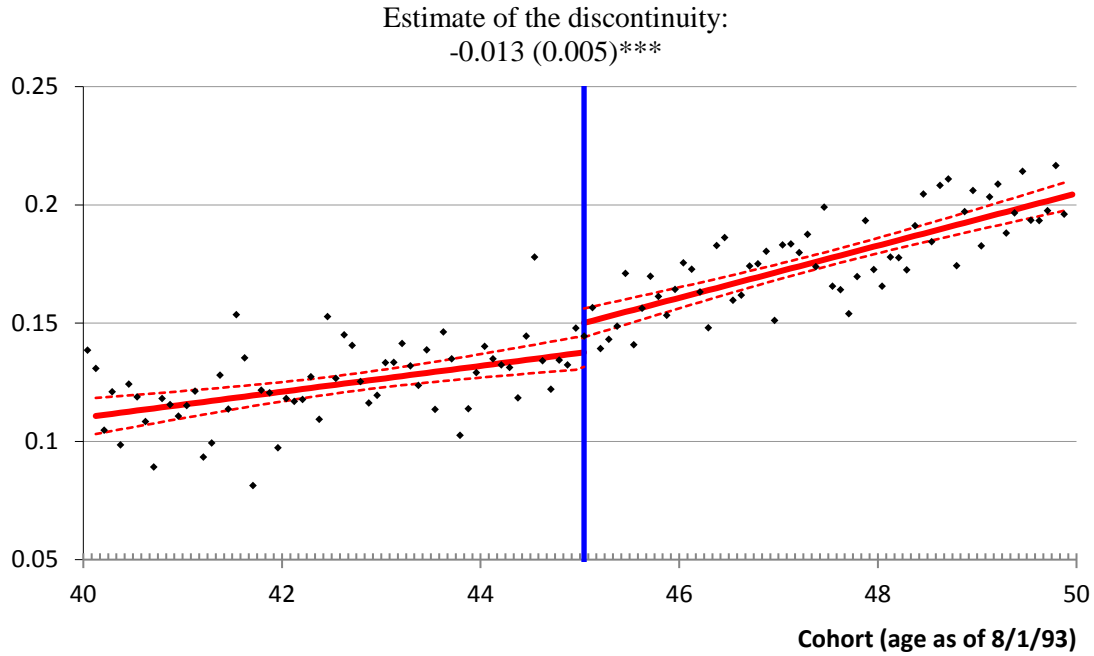
Estimate of the discontinuity:
-0.069 (0.007)



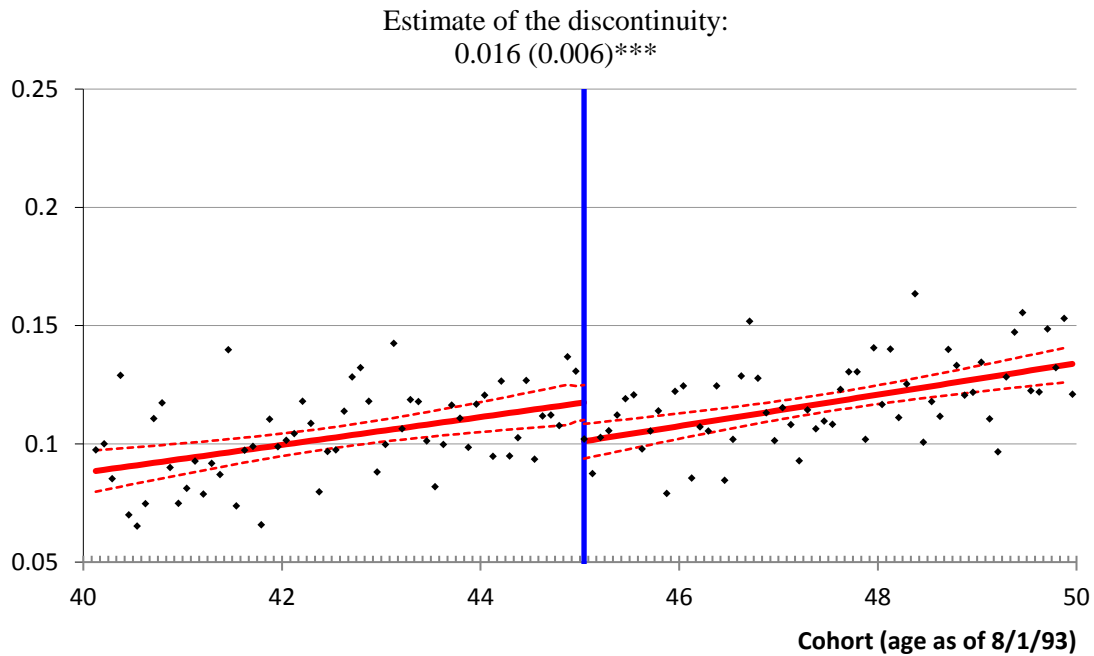
Note: The horizontal axis shows age of the DI claimant as of 8/1/93. Dashed lines represent the 95% confidence intervals. Regression estimates come from RD reduced-form regressions with the following controls: a dummy for being subject to the strict re-examinations (treatment), its interaction with [age-45], and age in months as of 8/1/93. Outcomes are studied in 1999 as this is the first year in which all re-examinations were finished.

Figure 2: Reduced-form effect of DI reform on mortality by 2010

A. Males



B. Females



Note: The horizontal axis shows age of the DI claimant as of 8/1/93. Dashed lines represent the 95% confidence intervals. Regression estimates come from RD reduced-form regressions with the following controls: a dummy for being subject to the strict re-examinations (treatment), its interaction with [age-45], and age in months as of 8/1/93. Similar graphs with quadratic age controls (not presented here) show a similar pattern, although the male effect drops in significance.

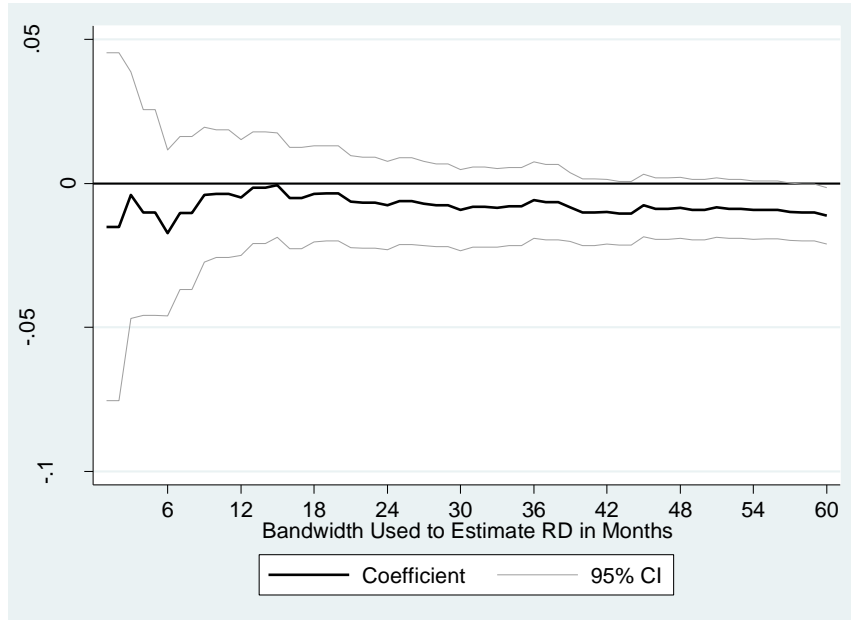
Figure 3: Reduced-form effects of DI Reform on mortality 1996-2010



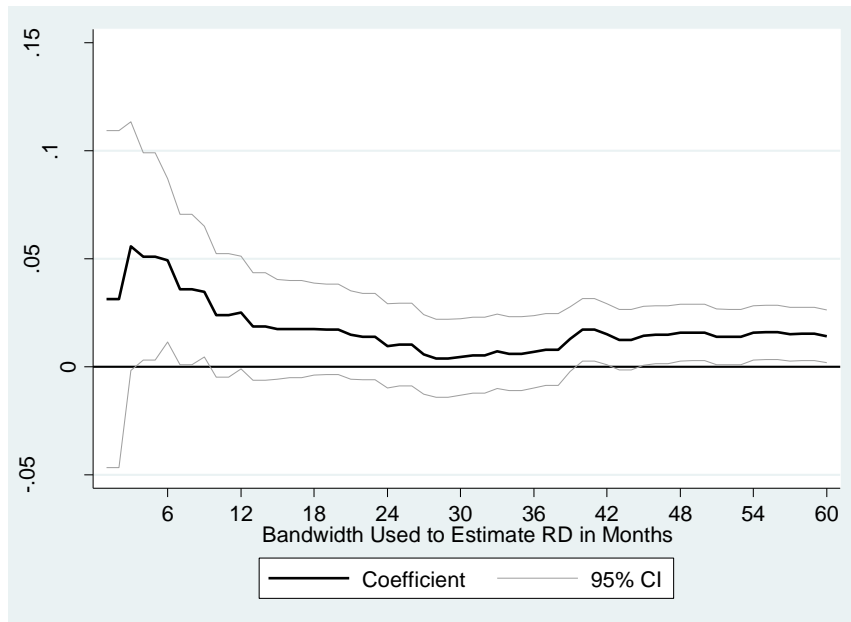
Note: Each marker in the figure represents a coefficient from a separate regression, in which mortality by a given year (see horizontal axis) is the dependent variable. Unfilled markers refer to coefficients not significantly different from zero at a 10% level. Black markers are significant at a 5% level, and grey markers are significant at the 10% level. Each dot in the figure comes from a separate regression. The following controls are used in the regressions: a dummy for being subject to the strict re-examinations (treatment), its interaction with [age-45], age in months as of 8/1/93, 6 dummies for degree of disability in 1996, pre-DI earnings, a dummy for being married in 1996, a dummy for being conceived during and born shortly after WWII (May 1945 – February 1946), a dummy for being conceived and born after WWII (February 1946 – January 1947), 39 regional dummies, 11 month of birth dummies, and duration in DI at the start of the reform.

Figure 4: Sensitivity to choice of bandwidth - Reduced-form effect on mortality by 2010.

A. Males

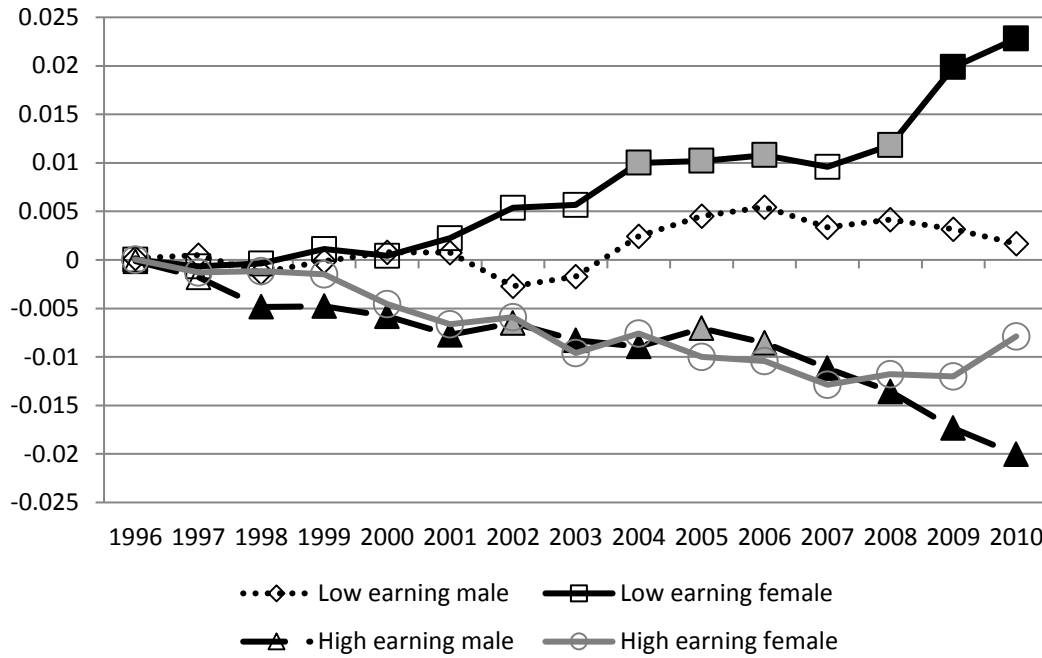


B. Females



Note: Estimates come from a set of RD reduced-form regressions in which the bandwidth is increased gradually by one-month from 1 to 60 months. The following controls are used in the regressions: a dummy for being subject to the strict re-examinations (treatment), its interaction with [age-45], age in months as of 8/1/93, 6 dummies for degree of disability in 1996, pre-DI earnings, a dummy for being married in 1996, a dummy for being conceived during and born shortly after WWII (May 1945 – February 1946), a dummy for being conceived and born after WWII (February 1946 – January 1947), 39 regional dummies, and duration in DI at the start of the reform.

Figure 5: Reduced-form effects of DI Reform on mortality 1996-2010, by previous earnings



Note: Each marker in the figure represents a coefficient from a separate regression, in which mortality by a given year (see horizontal axis) is the dependent variable. Unfilled markers refer to coefficients not significantly different from zero at a 10% level. Black markers are significant at a 5% level, and grey markers are significant at the 10% level. Estimates for each year come from a separate regression. The following controls are used in the regressions: a dummy for being subject to the strict re-examinations (treatment), its interaction with each of the four subgroups, its interaction with [age-45], age in months as of 8/1/93, a different age trend for each of the four

Table 1: Relation between Degree of Disability and Replacement Rates

Degree of disability:	Replacement rate (% of last earned wage):
80 – 100 %	70 %
65 – 80 %	50.75 %
55 – 65 %	42 %
45 – 55 %	35 %
35 – 45 %	28 %
25 – 35 %	21 %
15 – 25 %	14 %
Less than 15 %	0 %

Source: UWV (2006). UWV is the abbreviation of the agency that administers all social insurance for employees in the Netherlands. See text for a description of how the degree of disability is determined. Disability insurance benefit levels are determined as a percentage of the last earned wage and adjusted for inflation over time.

Table 2: Descriptive Statistics

	Males			Females		
	All	40.0-44.9	45.0-49.9	All	40.0-44.9	45.0-49.9
Panel A: Sample characteristics, measured prior to re-examination						
Male (dummy)	1	1	1	0	0	0
Marriage on Jan 1 st , 1996 (dummy)	69.32	66.18	71.53	60.37	57.78	62.21
Age on August 1 st , 1993	45.49	42.67	47.46	45.47	42.59	47.52
Duration on DI on August 1 st , 1993 (months)	101.38	96.78	104.60	90.28	91.14	89.67
Pre-DI earnings	7315.3	6987.4	7545.2	5046.2	5150.7	4972.0
DI benefits in 1999	11786.9	10023.4	13030.6	9028.7	8559.7	9362.8
Decrease in DI benefits (dummy)	7.02	12.20	3.36	7.00	13.28	2.53
Degree of disability (% of earnings capacity lost):						
15-25	7.4	9.0	6.4	4.8	5.5	4.4
25-35	11.5	13.2	10.3	5.0	5.7	4.5
35-45	9.0	9.8	8.5	3.5	3.9	3.2
45-55	7.1	6.5	7.4	4.9	4.9	4.9
55-65	2.4	2.1	2.6	1.6	1.7	1.5
65-80	2.5	2.2	2.6	1.2	1.2	1.2
80-100	60.1	57.1	62.2	79.0	77.1	80.4
Panel B: Mortality after re-examination						
<i>Mortality (%):</i>						
By 2010	15.6	12.5	17.7	11.2	10.3	11.8
N	91089	37547	53542	46670	19385	27285

Note: Since we have information available from 1996 onwards, the degree of disability is recorded in January 1996 (before the re-examinations).

Table 3: IV estimates. Effect of reform per 1000 €/year decrease in amount of DI benefits. Results for mortality 1999-2010

	Males	Females
1999	<0.001 (0.001)	0.003 (0.003)
2000	<0.001 (0.001)	0.002 (0.004)
2001	-0.001 (0.002)	0.002 (0.005)
2002	-0.002 (0.002)	0.005 (0.005)
2003	-0.002 (0.002)	0.004 (0.006)
2004	0.001 (0.002)	0.009 (0.007)
2005	<0.001 (0.003)	0.009 (0.007)
2006	<0.001 (0.003)	0.010 (0.008)
2007	-0.002 (0.003)	0.009 (0.008)
2008	-0.003 (0.003)	0.012 (0.009)
2009	-0.005 (0.003)*	0.021 (0.009)**
2010	-0.007 (0.003)*	0.024 (0.010)***
N	89618	46135

Note: Standard errors are in parentheses. Significance levels: * 10 percent; ** 5 percent; *** 1 percent. Each entry in the table comes from a separate regression. In the IV regressions the variable that is instrumented is the amount of DI, so the coefficients can be interpreted as effect size per €1000/year decrease in DI. The instrument itself is the treatment dummy (age less than 45 as of 8/1/93). The following controls are used in the regressions: a dummy for being subject to the strict re-examinations (treatment), its interaction with [age-45], age in months as of 8/1/93, 6 dummies for degree of disability in 1996, pre-DI earnings, a dummy for being married in 1996, a dummy for being conceived during and born shortly after WWII (May 1945 – February 1946), a dummy for being conceived and born after WWII (February 1946 – January 1947), 39 regional dummies, 11 month of birth dummies, and duration in DI at the start of the reform.

Table 4: Robustness and placebo tests (reduced-form results)

	Males	Females
Baseline results		
Mortality by 2010	-0.012 (0.005)**	0.014 (0.006)**
N	91089	46670
Panel A: No controls		
Mortality by 2010	-0.012 (0.005)**	0.014 (0.006)**
N	91089	46670
Panel B: Quadratic age trend		
Mortality by 2010	-0.012 (0.005)**	0.014 (0.006)**
N	91089	46670
Panel C: Sample with immigrants		
Mortality by 2010	-0.006 (0.005)	0.012 (0.005)**
N	108271	56872
Panel D: Placebo treatment (age cutoff at 55)		
Mortality by 2010	-0.00002 (0.005)	0.002 (0.006)
N	138763	62955
Panel E: Placebo sample (sample non-DI claimants)		
Mortality by 2010	0.0006 (0.0011)	-0.0004 (0.0010)
N	880091	895848

Note: Standard errors are in parentheses. Significance levels: * 10 percent; ** 5 percent; *** 1 percent. Each entry in the table comes from a separate regression. See the note to Table 3 for the demographic controls included in the regression. Estimates in Panel C also include 10 national origin dummies. Estimates in Panel E are performed on a sample of individuals aged 50-59 at the time of the reform.

Table 5: Results for different scaling variables

	Males	Females
Panel A: Instrument employment dummy (IV)		
First stage	0.037 (0.005)***	0.028 (0.006)***
	Effect of being re-employed in 1999 after the re-examination	
Mortality by 2010	-0.250 (0.135)*	0.533 (0.241)**
N	89618	46135
Panel B: Instrument total income (in €1000) (IV)		
First stage	-0.355 (0.155)**	-0.012 (0.131)
	Effect of 1000€/year decrease in total income in 1999	
Mortality by 2010	-0.033 (0.021)	
N	89618	

Note: Standard errors are in parentheses. Significance levels: * 10 percent; ** 5 percent; *** 1 percent. Each entry in the table comes from a separate regression. See the note to Table 3 for the demographic controls included in the regression.

Table 6: Different heterogeneous effects. Reduced-form effects of DI Reform

Mortality by 2010

Panel A: By DI-level in January 1996

$\beta_{\text{Full, Male}}$	-0.017 (0.006)***
$\beta_{\text{Full, Female}}$	0.014 (0.007)**
$\beta_{\text{Partial, Male}}$	-0.006 (0.007)
$\beta_{\text{Partial, Female}}$	0.022 (0.012)*
$\beta_{\text{Full, Male}} = \beta_{\text{Full, Female}}$	[<0.001]
$\beta_{\text{Partial, Male}} = \beta_{\text{Partial, Female}}$	[0.025]
$\beta_{\text{Full, Male}} = \beta_{\text{Partial, Male}}$	[0.210]
$\beta_{\text{Full, Female}} = \beta_{\text{Partial, Female}}$	[0.522]

Panel B: By time in DI at August 1st, 1993

$\beta_{\geq 5\text{yr, Male}}$	-0.020 (0.005)***
$\beta_{\geq 5\text{yr, Female}}$	0.022 (0.008)***
$\beta_{< 5\text{yr, Male}}$	<0.001 (0.007)
$\beta_{< 5\text{yr, Female}}$	0.008 (0.008)
$\beta_{\geq 5\text{yr, Male}} = \beta_{\geq 5\text{yr, Female}}$	[<0.001]
$\beta_{< 5\text{yr, Male}} = \beta_{< 5\text{yr, Female}}$	[0.429]
$\beta_{\geq 5\text{yr, Male}} = \beta_{< 5\text{yr, Male}}$	[0.006]
$\beta_{\geq 5\text{yr, Female}} = \beta_{< 5\text{yr, Female}}$	[0.179]

Panel C: By earnings before DI

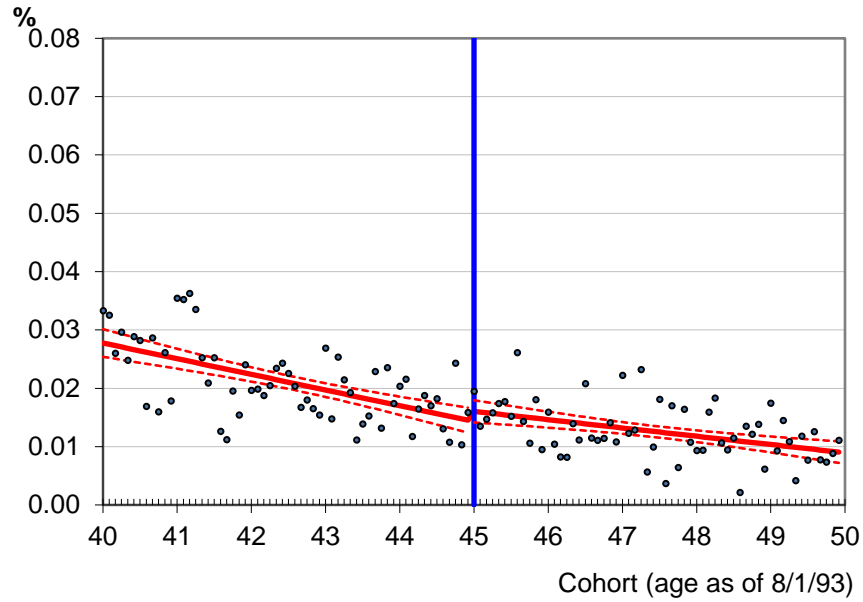
$\beta_{\text{Low, Male}}$	0.002 (0.007)
$\beta_{\text{Low, Female}}$	0.023 (0.007)***
$\beta_{\text{High, Male}}$	-0.020 (0.005)***
$\beta_{\text{High, Female}}$	-0.007 (0.011)
$\beta_{\text{Low, Male}} = \beta_{\text{Low, Female}}$	[0.024]
$\beta_{\text{High, Male}} = \beta_{\text{High, Female}}$	[0.645]
$\beta_{\text{Low, Male}} = \beta_{\text{High, Male}}$	[0.003]
$\beta_{\text{Low, Female}} = \beta_{\text{High, Female}}$	[0.004]

Note: Reduced form results. Standard errors are in parentheses. P-values in square brackets. Significance levels: * 10 percent; ** 5 percent; *** 1 percent. Each set of estimates per outcome and panel in the table comes from a separate regression. The following controls are used in the regressions: a dummy for being subject to the strict re-examinations (treatment), its interaction with each of the four subgroups, its interaction with [age-45], age in months as of 8/1/93, a different age trend for each of the four subgroups, 6 dummies for degree of disability in 1996, pre-DI earnings, a dummy for being married in 1996, a dummy for being conceived during and born shortly after WWII (May 1945 – February 1946), a dummy for being conceived and born after WWII (February 1946 – January 1947), gender, 39 regional dummies, 11 month of birth dummies, and duration in DI at the start of the reform.

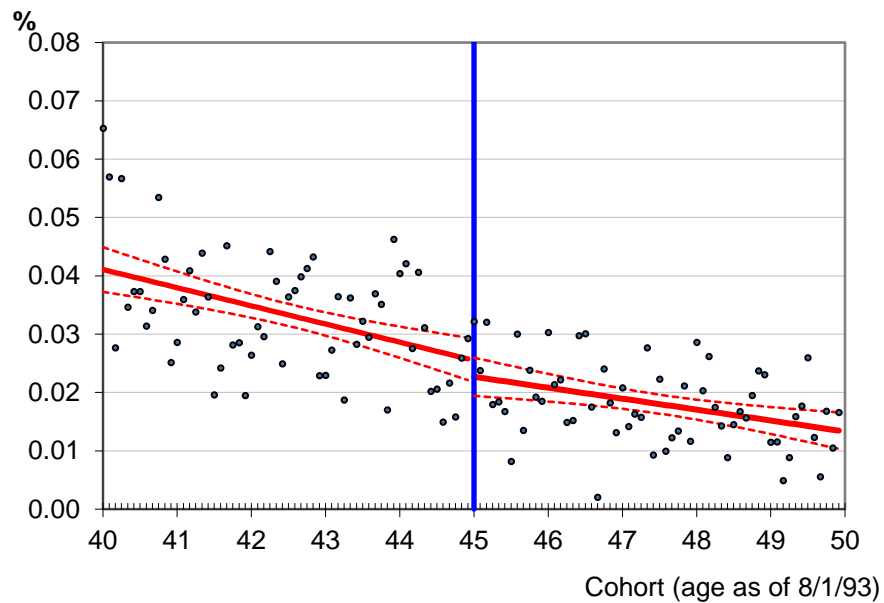
Appendix A: Additional figures and tables

Figure A1: Exit from DI in 1995

A. Males

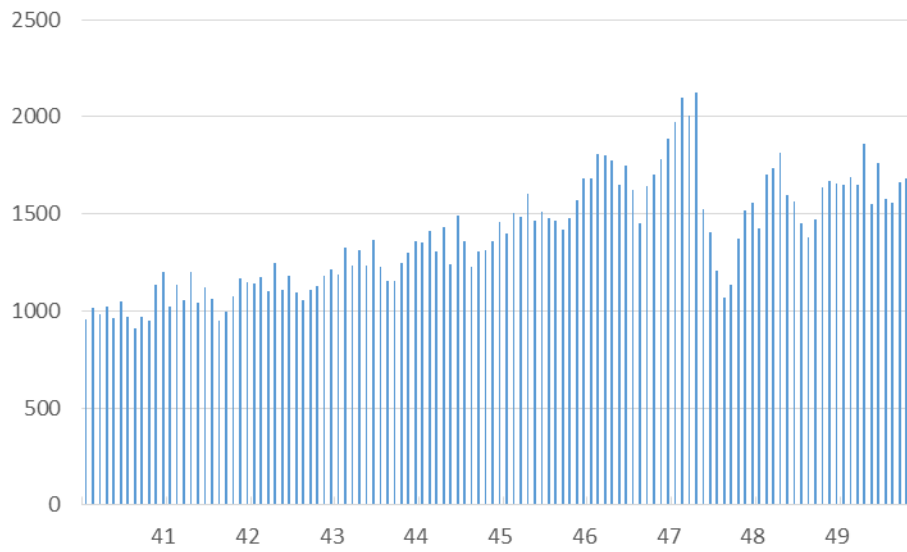


B. Females



Note: Immigrants are not excluded from this Figure as we cannot identify the country of origin for those who left DI in 1995. Regression estimates come from RD reduced-form regressions with the following controls: a dummy for being subject to the strict re-examinations (treatment), its interaction with [age-45], age in months as of 8/1/93.

Figure A2: Number of observations by cohort



Note: The McCrary density test is insignificant (p-value 0.114).

Table A1. Placebo tests. Reduced-form results for the effect of DI reform on different characteristics

Variable	Effect of the reform		Variable	Effect of the reform	
<i>Degree of disability</i>			<i>Corop region dummies</i>		
Dummy(15-25%)	0.003	(0.003)	Corop1	-0.002	(0.001)
Dummy(25-35%)	-0.005	(0.003)	Corop2	-0.001	(0.001)**
Dummy(35-45%)	-0.001	(0.003)	Corop3	0.000	(0.002)
Dummy(45-55%)	-0.002	(0.003)	Corop4	0.001	(0.002)
Dummy(55-65%)	0.002	(0.002)	Corop5	-0.001	(0.001)
Dummy(65-80%)	0.002	(0.002)	Corop6	0.001	(0.001)
Dummy(80-100%)	0.001	(0.005)	Corop7	0.000	(0.001)
Previous labor earnings	31.788	(38.017)	Corop8	0.001	(0.001)
No prev. lab. earn. info.	-0.003	(0.003)	Corop9	0.000	(0.001)
Dummy(married)	0.002	(0.005)	Corop10	-0.002	(0.002)
Duration in DI	0.594	(0.849)	Corop11	-0.002	(0.001)
			Corop12	0.001	(0.002)
			Corop13	-0.003	(0.002)
			Corop14	0.001	(0.002)
			Corop16	-0.001	(0.001)
			Corop17	0.000	(0.003)
			Corop18	-0.001	(0.002)
			Corop19	-0.002	(0.001)
			Corop20	-0.003	(0.001)**
			Corop21	-0.001	(0.001)
			Corop22	-0.002	(0.001)
			Corop23	-0.003	(0.003)
			Corop24	-0.002	(0.001)
			Corop25	0.000	(0.001)
			Corop26	-0.002	(0.002)
			Corop27	0.000	(0.001)
			Corop28	0.000	(0.001)
			Corop29	-0.003	(0.003)
			Corop30	0.000	(0.002)
			Corop31	0.001	(0.001)
			Corop32	0.001	(0.001)
			Corop33	0.000	(0.002)
			Corop34	0.002	(0.002)
			Corop35	0.002	(0.002)
			Corop36	0.004	(0.002)*
			Corop37	0.002	(0.001)
			Corop38	0.002	(0.001)
			Corop39	0.008	(0.003)***
			Corop40	0.001	(0.002)
N	137759				

Note: Standard errors are in parentheses. Significance levels: * 10 percent; ** 5 percent; *** 1 percent. Each entry in the table comes from a separate regression. See the note to Table 3 for the demographic controls included in the regression.

Table A2: Reduced-form results for the effect of DI reform on mortality 1996-2010

	Males		Females	
1996	<0.001	(<0.001)	<0.001	(<0.001)
1997	<0.001	(0.001)	-0.001	(0.001)
1998	-0.003	(0.002)*	-0.001	(0.002)
1999	-0.003	(0.002)	<0.001	(0.003)
2000	-0.004	(0.003)	<0.001	(0.003)
2001	-0.005	-0.003)	<0.001	(0.004)
2002	-0.005	(0.003)	0.002	(0.004)
2003	-0.006	(0.003)*	0.001	(0.004)
2004	-0.005	(0.004)	0.005	(0.005)
2005	-0.003	(0.004)	0.004	(0.005)
2006	-0.003	(0.004)	0.005	(0.005)
2007	-0.006	(0.004)	0.004	(0.005)
2008	-0.008	(0.005)	0.006	(0.006)
2009	-0.010	(0.005)**	0.012	(0.006)**
2010	-0.012	(0.005)**	0.014	(0.006)**
N	91089		46670	

Note: Standard errors are in parentheses. Significance levels: * 10 percent; ** 5 percent; *** 1 percent. Each entry in the table comes from a separate regression. See the note to Table 3 for the demographic controls included in the regression.

Table A3: Characteristics of compliers

	“Complier” characteristics	Average characteristics of the age 45.0 cohort	Difference between compliers and the rest of the sample for the age 45.0 cohort
Panel A: Males			
Degree DI in 1996 (scale of 1-8)	5.381 (0.402)***	5.315 (0.015)***	0.066 (0.401)
Degree DI in [15, 25)% in 1996	0.162 (0.055)***	0.074 (0.002)***	0.089 (0.055)
Degree DI in [25, 35)% in 1996	0.039 (0.068)	0.116 (0.002)***	-0.077 (0.067)
Degree DI in [35, 45)% in 1996	0.025 (0.058)	0.095 (0.002)***	-0.070 (0.058)
Degree DI in [45, 55)% in 1996	0.058 (0.052)	0.072 (0.002)***	-0.014 (0.052)
Degree DI in [55, 65)% in 1996	0.042 (0.028)	0.022 (0.001)***	0.020 (0.280)
Degree DI in [65, 80)% in 1996	0.092 (0.026)***	0.025 (0.001)***	0.067 (0.026)**
Degree DI in [80, 100)% in 1996	0.581 (0.078)***	0.597 (0.003)***	-0.016 (0.078)
Earnings prior to DI (Euro/yr)	6486 (544)***	7256 (22)***	-770 (542)
Dummy for missing pre-DI earnings	0.021 (0.051)	0.099 (0.002)	-0.078 (0.051)
Married	0.799 (0.065)	0.686 (0.003)***	0.113 (0.064)*
Duration on DI (months; as of 8/1993)	111.685 (10.993)***	102.577 (0.479)***	9.108 (10.928)
Panel B: Females			
Degree DI in 1996 (scale of 1-8)	4.985 (0.296)***	6.112 (0.0168)***	-1.127 (0.293)***
Degree DI in [15, 25)% in 1996	0.200 (0.043)***	0.051 (0.002)***	0.149 (0.043)***
Degree DI in [25, 35)% in 1996	0.078 (0.043)*	0.049 (0.002)***	0.029 (0.043)
Degree DI in [35, 45)% in 1996	0.085 (0.035)**	0.036 (0.002)***	0.049 (0.034)
Degree DI in [45, 55)% in 1996	0.001 (0.044)	0.049 (0.002)***	-0.048 (0.044)
Degree DI in [55, 65)% in 1996	0.042 (0.022)*	0.017 (0.001)***	0.025 (0.022)
Degree DI in [65, 80)% in 1996	-0.007 (0.014)	0.01 (0.001)***	-0.017 (0.014)
Degree DI in [80, 100)% in 1996	0.6 (0.062)***	0.788 (0.004)***	-0.188 (0.062)***
Earnings prior to DI (Euro/yr)	4349 (384)***	4998 (26)***	-648 (383)*
Dummy for missing pre-DI earnings	0.006 (0.026)	0.02 (0.001)***	-0.013 (0.025)
Married	0.707 (0.062)***	0.593 (0.004)	0.114 (0.062)*
Duration on DI (months; as of 8/1993)	88.801 (8.640)***	90.856 (0.693)***	-2.055 (8.601)

Note: Standard errors are in parentheses. Significance levels: * 10 percent; ** 5 percent; *** 1 percent. Effects for Corop regions are not presented in the Table, as there was hardly any effect of Corop regions on complier characteristics. “Compliers” are defined as individuals who are affected on the margin by the stricter re-examination. Hence, compliers refers to those who experience a decrease in their degree of disability following the stricter re-examination criteria but would not have experienced a decrease in their degree of disability under the old and more lenient criteria. The characteristics of the compliers are measured around the discontinuity, i.e. at the age 45.0 cohort. We follow the procedure described by Borghans et al. (2014) to infer the demographic characteristics of compliers. Standard errors are calculated by bootstrapping (5000 draws). N=91089 for males and N=46670 for females.