

## Social Support Substitution and the Earnings Rebound: Evidence from a Regression Discontinuity in Disability Insurance Reform<sup>†</sup>

By LEX BORGHANS, ANNE C. GIELEN, AND ERZO F. P. LUTTMER\*

*We exploit a cohort discontinuity in the stringency of Dutch disability reforms to estimate the effects of decreased DI (disability insurance) generosity on behavior of existing recipients. We find evidence of social support substitution: individuals on average offset €1.00 of lost DI benefits by collecting €0.30 more from other social assistance programs, but this benefit-substitution effect declines over time. Individuals also exhibit a rebound in earnings: earnings increase by €0.62 on average per euro of lost DI benefits and this effect remains roughly constant over time. This is strong evidence of substantial remaining earnings capacity among long-term claimants of DI. (JEL I38, J14, J22, J28, J31)*

**B**ecause estimates of labor supply responses are of tremendous policy relevance, the literature on the effects of changes in the eligibility criteria or generosity of social assistance programs<sup>1</sup> has rightly focused on labor supply responses. These estimates, however, do not capture two additional policy-relevant dimensions of the response to changes in social assistance programs. First, they miss potential spillover effects to other social assistance programs that arise when individuals substitute between programs. Such social support substitution reduces the welfare impact of reductions in generosity of any given program on recipients of that program. Moreover, social support substitution may decrease the reduced-form labor

\*Borghans: Department of Economics and ROA, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands (e-mail: Lex.Borghans@maastrichtuniversity.nl); Gielen: Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, and IZA (e-mail: Gielen@ese.eur.nl); Luttmer: Department of Economics, Dartmouth College, 6106 Rockefeller Center, Hanover, NH 03755 (e-mail: Erzo.FP.Luttmer@dartmouth.edu). We would like to thank Gerard van den Berg, Amy Finkelstein, Paul Frijters, Pierre Gielen, Ed Glaeser, Bart Golsteyn, David Johnston, Pierre Koning, Peter Kuhn, Ellen Meara, Jon Skinner, Doug Staiger, Frank Vella, and two anonymous referees for helpful comments. We thank seminar participants at IZA, Tilburg University, Maastricht University, University of Mannheim, University of Milan, the NBER Public Economics meetings, SciencesPo, University of New South Wales, Queensland University, the CPB, Erasmus University of Rotterdam, the University of Chicago, and the Tinbergen Institute, as well as participants at the IZA/IFAU, LEW, EALE, IZA/CEPR, and IZA/SOLE conferences, for insightful comments. We are grateful to Saaid Arshad, Meili Eubank, and Alexander Shen for providing excellent research assistance. We would like to especially thank Bas ter Weel, who greatly contributed to the conception of this paper. Statistics Netherlands has provided access to the data that was used in this project through a remote connection facility. As part of the data agreement, Statistics Netherlands has the right to review the results of our project prior to their dissemination to ensure that the confidentiality of the data is not unintentionally compromised and individual-specific information is not revealed. All errors are our own.

<sup>†</sup>Go to <http://dx.doi.org/10.1257/pol.6.4.34> to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

<sup>1</sup>We use the term *social assistance* generically to refer to any social insurance or income maintenance programs rather than to a specific program.

supply response to changes in generosity of a particular program when individuals take up other programs instead of adjusting their labor supply. Evidence on the extent of social support substitution is also important for policymakers because it allows them to make more accurate predictions of the budgetary impact of a reform to a social assistance program by taking into account the spillover effects of the reform on participation in other programs. Second, existing estimates of supply responses typically do not distinguish between responses by existing claimants and responses by (potential) new enrollees. Evidence of labor supply responses among existing DI recipients is of importance for policy reforms because effects which operate on the existing stock of recipients have the potential to make a much greater immediate impact than effects which operate on the comparatively small inflow into a DI program.

The Dutch DI system, which also insures against partial loss of earnings capacity, was significantly reformed in 1993. Two features make this reform particularly suitable for studying substitution between different social assistance programs as well as labor supply responses among current recipients. First, we have administrative panel data on the universe of Dutch disability insurance claimants, including information on their future labor market earnings and their future income from all other government cash social assistance programs. These data allow us to track for a period of nearly a decade what happens to (former) disability insurance claimants in the wake of the reform. Second, the reform contains a cohort discontinuity: the reform was significantly more stringent and led to an average benefit reduction of an additional 10 percent for the cohort that turned 45 after August 1, 1993. Because we have each individual's month of birth, we exploit this discontinuity by comparing later labor market earnings and social assistance income for the cohort just below this age cutoff to outcomes for the cohort just above the age cutoff. We scale this difference in outcomes by the discontinuity in disability benefit levels around the age cutoff. This yields two key ratios: (i) the benefit-substitution ratio, which is the average causal effect of the more stringent DI rules on income from other social assistance as a fraction of average lost DI income, and (ii) the earnings crowd-out ratio, which is the average causal effect of the more stringent DI rules on earnings as a fraction of average lost DI income.

We find that, in the short term (about two years after implementation of the reform), the more stringent DI rules increase the probability of receiving income from other social assistance programs by about 4.5 percentage points (on a base of 14 percent), and the income from these other social assistance programs replaces 30 percent of lost DI income. In other words, we find a substantial amount of social support substitution with a short-term benefit-substitution ratio of 0.30. The more stringent rules increase the probability of having any earnings by three percentage points (on a base of 35 percent) and also increase earnings on average. The additional earnings replace 62 percent of foregone DI income, yielding an earnings crowd-out ratio of 0.62. Recipients classified as fully disabled have a crowd-out ratio of 0.52, which is important evidence of a substantial labor supply response among long-term fully disabled individuals. Combining the effects of social assistance substitution and earnings crowd-out, we find that individuals are able to replace basically all of their foregone DI income on average. Specifically, we cannot reject that individuals

on average fully offset the cut in their DI benefit by increasing income from other social assistance programs and increasing earnings.

Our data allow us to follow the effects of the cohort discontinuity in the stringency of the reforms over a seven-year period, from approximately two to eight years after implementation of the reforms. Over this period, the benefit-substitution ratio falls from 0.30 to 0.14, though even after eight years the benefit-substitution ratio is still statistically significant. The earnings crowd-out ratio remains roughly constant over this period. Spillovers between social assistance programs may operate not only through former DI recipients' own choices, but could potentially also operate through the decisions of their spouses. However, we do not find any statistically significant evidence of responses by spouses in terms of labor supply or social assistance receipt. The point estimates indicate that spouses increase their labor supply but do not change their social assistance receipt. This implies that if we measure earnings crowd-out and benefit substitution at the household level rather than at the individual level, we find a higher earnings crowd-out ratio (about 18 percentage points higher) and a similar benefit-substitution ratio.

While the precise magnitudes of our findings are specific to this particular Dutch disability insurance reform, we believe our paper offers important lessons that are widely applicable. First, our evidence demonstrates that social support substitution occurs at an economically meaningful scale for prime-age disability insurance recipients. Hence, a carefully designed reform of a social assistance program would be well advised to take into account its effects on other social assistance programs. Second, our findings show that even long-term disability insurance recipients can still exhibit a meaningful rebound in their labor supply. Third, to measure the full impact of social insurance reforms on labor supply and reliance on other forms of social insurance, it is important to also consider effects over the longer term and to take possible behavioral responses of spouses into account.

Our findings on the existence of spillover effects between different social assistance programs confirm earlier results from other contexts.<sup>2</sup> With respect to child-related benefits, Garrett and Glied (2000) show that the increase in child Supplemental Security Income (SSI) eligibility in the early 1990s led to a greater increase in SSI enrollment in states with less generous benefits for Aid to Families with Dependent Children (AFDC). Moreover, Schmidt and Sevak (2004) show that states that reduced the generosity of their AFDC program experienced increases in SSI enrollments. Both studies suggest that families substitute between SSI and AFDC. Kubik (2003) shows that the substitution of SSI for AFDC is larger in states with negative fiscal shocks, suggesting that states actively encourage this

<sup>2</sup>An exception is a paper by Autor and Duggan (2008), in which they exploit a ruling that suddenly expanded the eligibility for Veterans' Disability Compensation (DC) for a subgroup of Vietnam Veterans. They find that the increased take-up of Veteran's Disability Compensation due to this ruling raised the receipt of Social Security Disability Insurance (SSDI) benefits. As Autor and Duggan note, this result may be explained by the fact that one needs leave the labor force to qualify for SSDI, and leaving the labor force is less costly for people who already receive DC. Thus, this institutional feature may explain the complementarity between two social assistance programs in this case. Another exception is Inderbitzin, Staubli, and Zweimüller (2013), who find that extended UI (unemployment insurance) benefits for workers in their early fifties increases the probability that these workers subsequently use Austria's relaxed DI systems for workers aged 55+ as a pathway to retirement. However, they do find that, for workers aged 55+, extended UI benefits substitute for DI.

substitution (because the state-share in SSI payments is generally lower than the state-share in AFDC payments). Duggan and Kearney (2007) examine panel data to find that households in which a child becomes eligible for SSI subsequently receive less income from AFDC, WIC, and food stamps.<sup>3</sup>

Regarding substitution among programs focusing on adults, Bound (1989) already noted that DI and other social assistance programs can act as substitutes, though he does not provide formal estimates of this effect. Duggan, Singleton, and Song (2007), Li and Maestas (2008), and Coe and Haverstick (2010) exploit differences by cohort in the generosity of Social Security retirement benefits to show that a reduction in these retirement benefits led to an increase in applications for or receipt of Social Security disability benefits. Koning and van Vuuren (2010) use Dutch administrative data to describe program enrollment after dismissal. They find that DI substitutes for UI but fail to find evidence that UI substitutes for DI. Using a regression discontinuity design, Lammers, Bloemen, and Hochguertel (2013) find that an increase in the stringency of job search requirements for older UI recipients in the Netherlands caused these recipients to be more likely to transition from UI to DI. Karlström, Palme, and Svensson (2008) use a difference-in-differences design to examine the effect of the abolition of DI as a path to early retirement for 60–64-year-olds in Sweden. They find that, in the two to three years following the reform, this group responded by taking up other forms of social assistance rather than by increasing their labor supply. Finally, Staubli (2011) also uses a difference-in-differences approach to show that a disability insurance reform that affected 55–56-year-old males in Austria had spillover effects on their take-up of unemployment insurance and sick leave, in addition to affecting their labor supply.

Our paper contributes to this literature by estimating substitution between social support programs for prime-age individuals in a setting that allows us to very cleanly identify the degree of spillovers between programs. In addition, we extend the literature by examining substitution effects over longer horizons (up to eight years after the reexaminations for our cohorts took place).

Our estimates also contribute to an extensive literature on the labor supply effects of disability insurance (see Bound and Burkhauser 1999 for an overview). Parsons (1980) presents cross-sectional evidence suggesting that the rise in DI generosity has contributed to the decline in male labor force participation. Gruber (2000) exploits a natural experiment in Canada and finds a sizeable labor force participation response from older workers to a change in the generosity of DI benefits. Much of the more recent work in the United States on labor supply effects of DI compares accepted to rejected DI applicants. Bound (1989), von Wachter, Song, and Manchester (2011), and Singleton (2012) compare accepted and rejected applicants directly, and those estimates are probably upper bounds of labor supply effects because there are likely unmeasured determinants of the rejection decision that are correlated with labor supply. To overcome this issue, other studies use plausibly exogenous variation in rejection rates. Gruber and Kubik (1997) use variation across states and time in rejection rates; Chen and van der Klaauw (2008) use an age discontinuity in rejection rates

<sup>3</sup>WIC (Women, Infants, and Children) provides nutritional assistance to low-income families with young children and pregnant women.

for a particular subgroup; De Jong, Lindeboom, and van der Klaauw (2011) use variation in screening severity induced by an experimental intervention; French and Song (2011) use variation in rejection rates due to the essentially random assignment of administrative law judges to DI cases; and Maestas, Mullen, and Strand (2013) use variation in rejection rates due to the essentially random assignment of DI examiners to DI cases. These studies all find clear evidence of labor supply responses to disability insurance. Autor and Duggan (2003) exploit the interaction of state disability replacement rates with national changes in program stringency to find credible evidence that increased DI generosity reduced labor force participation of high school dropouts.

Despite this extensive literature, much less is known about the degree to which existing long-term DI recipients can be induced to work. A number of recent papers have started to amass evidence on this topic. Maestas and Song (2011) use the automatic conversion of DI into Social Security benefits to show that there is a labor supply effect from DI on older disability insurance recipients. Campolieti and Riddell (2012) use a difference-in-differences design to show that an increase in the earnings exemption increased the propensity to work among existing DI recipients in Canada. Moore (2012) examines the labor supply response of existing DI recipients who lost eligibility due to a 1996 US reform that removed alcohol and drug addictions as disabling conditions. He finds a 22-percentage-point increase in employment as a result of these terminations, and notes that the employment response has an inverse-U shape with a maximum for those who have been on DI for about three years. Finally, Kostøl and Mogstad (2014) exploit a regression discontinuity to examine whether financial incentives can induce existing DI recipients in Norway to increase their labor supply. They find that prime-age DI recipients respond to these incentives but that DI recipients approaching retirement age do not show a significant response. Our study contributes to this emerging literature on the work capabilities of existing DI recipients; we provide well-identified estimates from a different setting to show strong labor supply responses to changes in DI generosity among prime-age long-term disability insurance recipients, including individuals classified as fully disabled.

### **I. The 1993 Dutch Disability Insurance Reform**

After a waiting period of one year, individuals in the Netherlands are entitled to disability benefits if an illness or infirmity prevents them from earning the amount they used to earn before the onset of the disability.<sup>4</sup> The replacement rate offered by DI depends on the “degree of disability,” which is defined by the percentage difference between the prior earnings and the remaining potential earnings capacity of the DI applicant. For readers familiar with the US disability system, it is useful to keep four key differences between the US and Dutch systems in mind. First, unlike the US system, the Dutch DI system insures against partial disability, and DI recipients

<sup>4</sup>Also see Bovenberg (2000), who provides useful institutional background information on the Dutch disability act. See García-Gómez, von Gaudecker, and Lindeboom (2011) for further background information and descriptive evidence of DI enrollment trends and patterns.

in the Netherlands can simultaneously work (for the fraction of “remaining earnings capacity”) and receive DI (for the fraction of “lost earnings capacity”). Second, health insurance and other benefits are not tied to receipt of DI. Third, DI benefits are only paid to the disabled worker and not adjusted for dependents. Fourth, the replacement rate only depends on the disability rating, and does not vary by work history or by level of earnings below the maximal covered earnings. Moreover, the replacement rate for a fully disabled worker is 70 percent, which is much higher than the cash replacement rates of between 40 percent and 50 percent that Autor and Duggan (2003) estimate for US workers with median earnings.

In order to explain the cohort discontinuity in the 1993 DI reform, we first describe how the Dutch disability insurance system determined eligibility and replacement rates prior to the reform. Before 1993, the remaining earnings capacity was determined by the following procedure. First, a medical doctor examined the applicant and compiled a list of work activities that, according to the doctor’s judgment, the applicant could still perform.<sup>5</sup> Second, using a dictionary of occupations that specified for each occupation the required education level and work activities, a list of occupations that an applicant could still perform was compiled, but occupations that were more than two “education levels” (on a seven-level scale) below the education level required for the applicant’s previous occupation were excluded. Finally, if the list contained at least five suitable occupations with at least ten active workers in the applicant’s region,<sup>6</sup> then the mean wage of the five highest-paying occupations on the list was taken as the applicant’s potential earnings capacity. The loss of earnings due to the disability, measured by the difference between the prior earnings and the potential earnings capacity, determined the individual’s disability rating, or “degree of disability.” If it was not possible to specify five suitable occupations with at least ten workers, the degree of disability was set at 100 percent. The measured disability ratings were grouped into eight categories varying from 0–15 percent to 80–100 percent, and these categories determined the replacement rate (see Table 1). The replacement rate was applied to the individual’s indexed previous earnings, where the previous earnings were subject to a cap (about €36,000/year in 1999). Individuals on DI had an earnings exemption equal to their capped indexed previous earnings times the degree they were not disabled (which was set at 100 percent minus the lower bound of their degree-of-disability category). If a DI beneficiary chose to work and had earnings above this exemption, the person’s DI benefits were reduced in the short term and disability rating adjusted in the longer term (typically after about three years).<sup>7</sup> Because adjustments due to earnings exceeding the

<sup>5</sup>The list includes 27 physical activities (such as “lifting,” “kneeling,” and “ability to deal with temperature fluctuations”) and 10 psychological abilities (such as “ability to work under time pressure,” “ability to perform monotonous work,” and “ability to deal with conflict”).

<sup>6</sup>The Netherlands was divided up into five regions and 16 “start regions.” Alternative occupations had to be found in the “start regions” first. The labor market expert could only look for occupations in neighboring regions (within one of the main five regions) if the start region contained fewer than five suitable occupations.

<sup>7</sup>In the short term, DI benefits are paid assuming the person’s degree of disability is reclassified based on his current earnings, but the reclassification does not formally happen because the higher level of earnings is not yet considered sufficient evidence of a lasting increase in earnings capacity. Instead, the actual reclassification only happens if the person has exceeded the earnings cap for three years, at which point the increase in earnings capacity is considered lasting. Because 1999 is less than three years after the reexaminations took place for the cohorts relevant to our analysis, the disability ratings in 1999 do not depend on endogenous reactions to the reexaminations.

TABLE 1—RELATION BETWEEN DEGREE OF DISABILITY AND REPLACEMENT RATES

Range for degree of disability (percentages)	Replacement rate (percent 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
[ 0, 15)	0

*Notes:* 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.

*Source:* UWV (2006). UWV is the abbreviation of the agency that administers all social insurance for employees in the Netherlands.

exemption amount took place independently of reexaminations, the anticipation of being reexamined provided no special incentive for individuals to adjust their earnings.

The DI reforms of 1993 tightened this procedure in two respects. First, the determination of disability had to be based on objective medical information. In other words, the applicant needed to have a clearly observable and functional work limitation, as well as a plausible and direct relationship between the functional work limitation and the medical diagnosis. Second, the criteria for the list of qualifying alternative occupations were relaxed: (i) occupations more than two education levels below the applicant's education level were included from 1993 on, (ii) the list only needed to contain three qualifying alternative occupations (rather than five), and (iii) the geographic region in which these occupations had to exist with at least ten active workers was expanded roughly threefold.<sup>8</sup> The remaining earnings capacity was henceforth determined by the three highest-paying occupations on the list. If these changes affected the remaining earnings capacity, they increased it, thereby decreasing the disability rating. The average of the highest-paying occupations must weakly rise as occupations were added the list. The average of the three highest-paying occupations is, by definition, weakly higher than the average of the five highest-paying occupations. Finally, the expanded list made it less likely that the list would not contain at least the minimum number (now three) of occupations needed to avoid declaring the applicant as fully disabled. As a result, for any given individual, the disability rating under the new criteria is always weakly lower than the disability rating this individual would have received under the old criteria.<sup>9</sup> This means that no individual would have received a higher replacement rate under the

<sup>8</sup>Now, all suitable occupations within the main region (out of five main regions) where the individual is residing are used to calculate the potential earnings capacity.

<sup>9</sup>Another important change of the 1993 DI reform was the introduction of an age- and duration-dependent benefit for new applicants. To those already receiving disability benefits as of August 1993 (i.e., the group that we are studying here), these changes did not apply and the benefit level remained a function of the indexed previous earnings.

new criteria than what this individual would have received if the old criteria had been used.

A reduction of the disability rating mechanically increased the earnings exemption. However, the net effect was still that, at any given level of earnings, DI benefits calculated under the new procedure were the same as or lower than DI benefits calculated under the old procedure. In other words, an individual's budget set under the new procedure was weakly dominated by the budget set faced by this same individual under the old procedure. It is important to note that the size of the benefit cut was not uniform across people, and we explore the heterogeneity in benefit reductions below.

The new procedure for determining benefits was applied to new DI applicants as well as to existing DI claimants who were 50 or younger at the time the reform went into effect. Because reexaminations of existing claimants are time-consuming, these reexaminations were scheduled to take place by cohort over a period of several years. Disability claimants who were age 34 or younger on the August 1, 1993 were reexamined in 1994, the 35–40-year-old cohort in 1995, the 41–44-year-old cohort in 1996–1997, and the 45–50-year-old cohort in 1997–2001. However, on November 12, 1996, a parliamentary motion was passed stipulating that the age 45–50 cohort would be reexamined based on the previous and more generous procedure for determining replacement rates. This motion generated a discontinuity between the under-45 cohorts and the 45+ cohorts in the generosity of DI, and we exploit this discontinuity to estimate the effects of a decrease in DI generosity on the behavior of existing DI recipients.<sup>10</sup> We note that we estimate the effect of a discontinuity in DI generosity in a context where everyone was reexamined (both the under-45 and the 45+ cohorts). If there is an interaction effect between a change in DI generosity and being reexamined, the effect of a change in DI generosity on labor supply and social support substitution would be different in a context without any reexaminations.

## II. Data

### A. Data Sources

This paper uses administrative data that Statistics Netherlands has assembled from several sources. These data are merged at the individual level by using a so-called RIN-number (a coded version of the Dutch equivalent of the US Social Security number).<sup>11</sup>

We have administrative data (AO) on all disability benefit recipients aged 15–64 in the Netherlands for the period 1995–2005. The data were collected by the organizations responsible for administering disability benefits and they contain the start and end dates of a disability spell, the disability rating (in categories), disability

<sup>10</sup>Because the discontinuity in the stringency of disability reform applies to existing recipients, we can estimate the effects of this discontinuity only for those who were receiving disability insurance at the time the reform went into effect. Our setting does not allow us to estimate effects stemming from people who would have flowed into DI under the less stringent rules but not under the more stringent rules.

<sup>11</sup>Researchers can use these data via a remote-access computer after signing a confidentiality agreement.

benefit payments, earnings prior to the DI spell, and the reason for the termination of the disability spell. However, the data do not contain reliable or consistent information about industry or about the medical condition that gave rise to the disability spell.

We obtained the demographic characteristics of the disability claimants from the municipal registries (GBA), which contain all residents of the Netherlands. This database includes information on each person's month and year of birth, marital status, number of children, national origin, and place of residence, as well as the identification numbers (RIN-codes) of their partners. The RIN-codes of the partners allow us to merge in data on income sources of the partner. We collected the demographic characteristics as of January 1996, which is the start of our sample period.

Finally, we obtained information on labor market earnings and sources of social assistance income other than DI by merging five administrative datasets: earnings of all employees, self-employment earnings, unemployment benefits (WW), general assistance (Bijstand), and receipt of any other form of social assistance (from about 30 relatively minor programs). Data on social assistance come from the organizations that administer these programs. Information about the earnings from paid labor and self-employment are gathered by Statistics Netherlands using information from the tax authorities and social insurance records. All these files are available from 1999 onward, which is why 1999 is the start year for our empirical analysis. Unemployment insurance covers any income loss due to unemployment for a duration of up to five years, where the duration depends on one's work history. General assistance is unlimited in duration and does not require dependents (unlike the US welfare program), but it is means tested. Apart from the programs mentioned here, there are no additional cash social assistance programs in the Netherlands that are relevant for individuals in the age range of our sample. Exact variable definitions are provided in online Appendix B.

### B. Sample Definition

In our baseline analysis, we restrict the sample to all individuals (i) who received disability benefits on August 1, 1993, (ii) who were between the ages of 42.5 and 47.5 at that date, and (iii) who were still on DI as of January 1, 1996. The first restriction is necessary because the discontinuity in benefit rules only applied to existing claimants on the date the reform went into effect. The second restriction limits the sample to cohorts close to the age where the discontinuity in benefit rules occurs. We selected the bandwidth of  $+/- 2.5$  years based on the criterion by Imbens and Kalyanaraman (2012).<sup>12</sup> The last restriction is driven by data availability. While our data on disability start in 1995, the information in the 1995 file does not contain earnings prior to DI, so we use the files from 1996 onward instead. Thus, we can only observe individuals who were on disability at the time the reform legislation was passed if they remained on disability until January of 1996 or later. We believe

<sup>12</sup>The Imbens-Kalyanaraman criterion yields different optimal bandwidths for different outcome variables. Rather than changing the sample for each outcome variable, we selected a bandwidth in the middle of the range suggested by the Imbens-Kalyanaraman criterion, and applied this bandwidth to all our specifications.

it highly unlikely that differential attrition occurred around the age discontinuity prior to January of 1996, as the reexaminations for the individuals in our sample did not start until later in 1996 and the government decided only in November of 1996 that the age 45+ cohort would not be subject to the new, more stringent criteria.<sup>13</sup> Hence, only after the start of our sample period did DI recipients become aware of the discontinuity in benefit generosity that we exploit in our analysis. A plot of the density of disability claimants by cohort is relatively smooth and further alleviates any concerns about differential attrition prior to the start of our sample period (see online Appendix Figure A1). The plot also indicates that heaping-induced bias (Barreca, Lindo, and Waddell 2011) is not a concern and that there is no discontinuity in the density around the cutoff age of 45 ( $p$ -value of 0.126 in the McCrary 2008 density test). We end our sample in 2005 because DI was fundamentally reformed in 2006.

We exclude all individuals who appeared on more than one disability record in our data in a given month (about 3 percent of the sample). We exclude these observations because it is not clear whether they reflect administrative/coding errors or whether they truly concern individuals who are entitled to multiple different disability insurance benefits. We have checked that no discontinuity occurs at age 45 in the likelihood that an individual has more than one disability record. Therefore, we are not concerned that the omission or inclusion of the 3 percent of observations with multiple records would substantively affect our results. After these sample restrictions, our baseline sample contains 84,185 observations.

### C. Summary Statistics

Table 2 presents summary statistics for our key variables. Panel A shows the characteristics of our sample measured as of January 1996 (before the reexaminations took place). About one-third of disability claimants are female and about two-thirds are married. The average DI spell started in 1985. Thus, when the reform was implemented in 1996–1998 for the cohorts in our sample, the average claimant in our data had been on DI for more than a decade. About two-thirds of the sample is classified as fully disabled (earnings capacity reduction of 80 percent or more) and is therefore eligible for a replacement rate of 70 percent. The fraction fully disabled is markedly higher among females than among males. Only about 4 percent of the sample is considered to have lost between 55 percent and 80 percent of their earnings capacity. The remaining 30 percent of the sample is considered to have lost between 15 percent and 55 percent of their earnings capacity and is eligible for replacement rates between 14 percent and 35 percent. The last two columns of Table 2 compare the under-45 cohort to the 45+ cohort. Their sample means are reasonably similar. Of course, our RD (regression discontinuity) design allows for smooth differences in characteristics by cohort. Only discontinuities in characteristics at the age cutoff would be problematic, and we will test for such discontinuities in the next section.

<sup>13</sup>Please note that we refer to cohorts by their age as of August 1, 1993.

TABLE 2—DESCRIPTIVE STATISTICS

	Full		Age		Age
	sample	Males	Females	42.50–44.99	45.00–47.49
<i>Panel A. Sample characteristics, measured prior to reexamination</i>					
Female (0=no; 1=yes)	0.34	0.00	1.00	0.34	0.34
Married (0=no; 1=yes)	0.66	0.69	0.61	0.65	0.67
Age on August 1, 1993	45.18	45.19	45.18	43.70	46.27
Start date of DI spell (year)	1985.1	1984.8	1985.7	1985.3	1985.0
Degree of disability (percent of earnings capacity lost):					
[15–25)	7.67	8.85	5.36	9.70	6.12
[25–35)	9.53	11.96	4.74	10.73	8.61
[35–45)	6.91	8.75	3.28	6.98	6.85
[45–55)	5.78	6.32	4.74	5.37	6.10
[55–65)	2.01	2.22	1.60	1.97	2.04
[65–80)	1.97	2.41	1.11	1.93	2.00
[80–100]	66.10	59.50	79.17	63.32	68.28
<i>Panel B. Outcomes after reexamination</i>					
Labor market status in 1999 (percent):					
Still on DI (on the original spell)	91.75	91.68	91.89	87.25	95.18
Employed	35.75	44.78	18.03	39.14	33.18
Reentry in DI (new spell)	0.23	0.20	0.30	0.37	0.13
Other social assistance	15.32	14.63	16.68	17.83	13.41
Zero income (dummy for no formal income)	3.91	3.66	4.39	4.24	3.66
Labor market status in 2005 (percent):					
Still on DI (on the original spell)	81.01	80.04	82.91	76.63	84.34
Employed	28.84	36.36	14.09	33.03	25.67
Reentry in DI (new spell)	4.62	4.33	5.18	5.50	3.95
Other social assistance	21.44	23.28	17.84	20.83	21.91
Zero income (dummy for no formal income)	8.46	8.97	7.46	8.39	8.51
Income by source in 1999, €/year (including zeros):					
DI from original DI spell	10,296	11,135	8,649	9,299	11,054
Earnings	5,916	7,753	2,309	6,500	5,471
New DI spell	25	22	31	39	15
Other social assistance	924	839	1,089	1,060	820
Income by source in 1999, €/year (if nonzero):					
DI from original DI spell	11,731	12,732	9,785	11,201	12,098
Earnings	17,045	17,814	13,282	17,118	16,979
New DI spell	10,715	11,105	10,206	10,568	11,027
Other social assistance	6,080	5,811	6,543	6,008	6,153
Income by source in 2005, €/year (including zeros):					
DI from original DI spell	11,421	12,343	9,611	10,351	12,235
Earnings	5,452	7,136	2,145	6,485	4,666
New DI spell	618	624	606	692	561
Other social assistance	1,322	1,231	1,501	1,233	1,389
Income by source in 2005, €/year (if nonzero):					
DI from original DI spell	14,491	15,887	11,862	13,920	14,884
Earnings	20,136	20,889	16,308	20,822	19,460
New DI spell	13,376	14,390	11,709	14,204	12,593
Other social assistance	6,185	5,307	8,433	5,951	6,353
Observations	84,185	55,772	28,413	36,362	47,823

*Notes:* Since we have information available from 1996 onward, both marital status and degree of disability are recorded in January 1996 (before the reexaminations). Age is measured as of August 1, 1993, so age is a measure for cohort.

Panel B presents the means of our key outcome variables. We only present the values for 1999 and 2005 in the interest of space. In 1999, about one to three years after the reexaminations took place for the age cohorts in our sample, 92 percent of those on DI at the start of 1996 were still on DI (where being on DI in 1999 is defined as having received positive income from DI in that year). About 36 percent of our sample was working, defined as having positive earnings (including those from self-employment) in 1999, which is consistent with DI also covering partial disability in the Netherlands. Of those who had left DI, 53 percent were employed, whereas 33 percent of those on DI were employed; thus, a considerable number combined DI receipt with work. The fraction of men working (45 percent) was more than twice as high as the fraction of females with positive earnings (18 percent). Fifteen percent of our sample also had social assistance income other than DI in 1999. Another 4 percent were not observed in any of our administrative files. Most of these individuals did not have any formal labor or social assistance income in 1999, but about one-third of them died or emigrated during our sample period.<sup>14</sup>

In 1999, the average income in our sample was about €17,000, of which roughly two-thirds came from DI benefits, with the remaining third coming mostly from earnings. Income from other social assistance programs accounted only for about 6 percent of total income. In 2005, about seven to nine years after the reexaminations, 81 percent of those on DI at the start of 1996 were still on DI. Between 1999 and 2005, the fraction employed fell from 36 percent to 29 percent, and the fraction with income from social assistance other than DI increased from 15 percent to 21 percent. These trends are consistent with the general decline in labor force participation in the Netherlands as people approach retirement. In 2005, about two-thirds of total income in our sample still came from DI benefits.

### III. Results

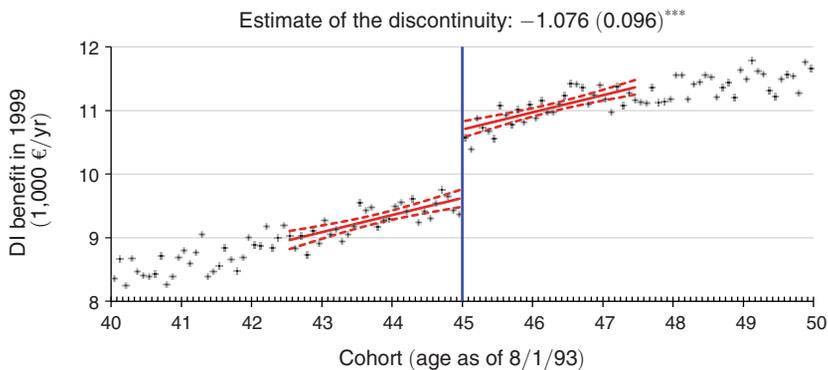
#### A. Magnitude of the Reform

To what extent did the more stringent reexaminations reduce the generosity of the DI program for the under-45 cohort? The answer to this question allows us to interpret the magnitude of the effects of the reform on earnings and on receipt of other forms of social assistance. Figure 1 shows three measures by which to gauge the magnitude of the reform: the effect on benefit amounts, the effect on replacement rates, and the effect on participation in the DI program.

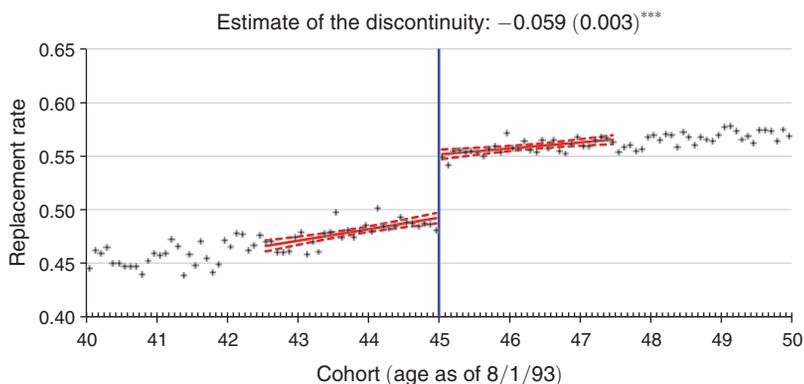
Panel A plots annual disability benefit amounts by cohort in 1999, including zeros for those who exited. Visually, there is a clear discontinuity at the cutoff age. We estimate the size of this discontinuity by running an OLS regression of the outcome variable on a treatment dummy that equals 1 for cohorts subject to the more stringent reexaminations (i.e., age less than 45), a linear term in age, and an interaction

<sup>14</sup>About 1.1 percent of our sample had died by 1999 and 0.3 percent had emigrated. These observations are included in the main analysis and their income and participation variables are all set to zero in 1999. Results are extremely similar if we exclude these observations altogether. The discontinuity in the stringency of the reexamination had no significant impact on the probability of death or the probability of emigration.

Panel A. Effect on DI benefit amount



Panel B. Effect on the DI replacement rate



Panel C. Effect on participation in DI

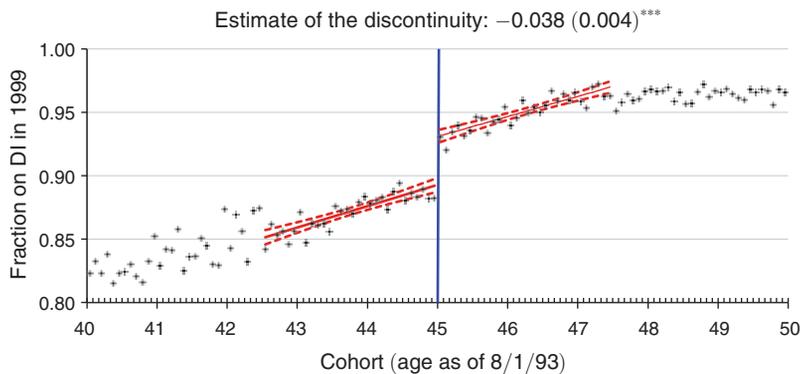


FIGURE 1. MAGNITUDE OF THE REFORM

*Notes:* Standard errors are in parentheses. Each figure is based on 84,185 observations. The dotted lines represent the 95 percent confidence intervals. Regression estimates come from reduced-form RD regressions without demographic control variables.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

of the treatment dummy with the variable [age–45]. All ages are specified as of August 1, 1993, so they effectively measure cohorts. We run this specification for all RD estimates reported in the figures in the paper. The fitted regression line is indicated in the plot, and the RD estimate on the treatment dummy is  $-1.076$ , indicating that the more stringent reexaminations for the younger cohort reduced their annual DI benefits by €1,076, or about 10 percent. All reduced-form RD estimates can be seen as local average treatment effects—local in the sense that they only apply at the discontinuity (i.e., for the cohort turning 45 on August 1, 1993) and average in the sense that it is the average effect for all those at the discontinuity. The key identifying assumption behind all of our RD estimates is that the only discontinuous change at the age cutoff is the stringency of the DI reexaminations. While we cannot test this assumption, we know of no other policy changes that would create a discontinuity at this cutoff. Further, when we run our reduced-form RD specification using all of our demographic characteristics as dependent variables, only 2 out of the 39 regression coefficients are significant.<sup>15</sup> Moreover, the RD coefficients in the placebo regressions are jointly insignificant ( $p$ -value: 0.4990).

Panel B of Figure 1 shows that the replacement rate, including zeros for those who exited, is 5.9 percentage points lower for the affected cohort at the discontinuity.<sup>16</sup> The average replacement rate for those who just escaped the more stringent reexaminations is 0.55; the 5.9-percentage-point drop therefore represents an 11 percent decline. Panel C shows that the fraction of the sample that is still on the original DI spell in 1999 falls discontinuously by 3.8 percentage points at the age cutoff. Overall, Figure 1 shows that the more stringent reexaminations roughly translate into a 10 percent benefit reduction, though it should be kept in mind that the size of the benefit cut was not uniform across individuals. The effects of the reform on labor supply and other benefit receipt should be viewed in light of this magnitude.

Given that the benefit cuts were not uniform, we want to determine whose benefits were reduced by the reform. In other words, we want to describe the demographic composition of those who would not experience a reduction in their disability rating (and, thus, benefits) under the less stringent reexamination, but do experience such a reduction under the more stringent reexamination. Following standard IV terminology, we refer to this group as the *compliers*.<sup>17</sup> Because the definition of compliers is based on a counterfactual outcome, the demographic characteristics of the compliers cannot simply be measured. However, as detailed in online Appendix A, we can

<sup>15</sup>The full results are presented in online Appendix Table A1. It would be instructive to do similar checks on the identification strategy with our key outcome variables: labor income and income from other social assistance programs. Unfortunately, we do not have data on these variables prior to 1999.

<sup>16</sup>The data do not contain the post-reform replacement rate for those who exited from the DI administration. Based on discussions with the DI administration and based on the economic incentives faced by beneficiaries, our understanding is that most exits occurred for those who were no longer eligible for DI. In particular, beneficiaries who had their benefits reduced to zero because of excess earnings could nevertheless stay in the DI administration with a positive disability rating (and, therefore, a nonzero nominal replacement rate) for about three years. Individuals had an incentive to do so because it gave them the option of receiving future DI benefits without having to requalify if their earnings would fall back. However, we cannot completely rule out that some of those who exited from the DI administration were still eligible for a positive replacement rate.

<sup>17</sup>Although the term *complier* is standard IV terminology, it is somewhat counterintuitive in our setting, as “compliance” in our setting is determined by the procedure that translates the results of the reexamination into benefit levels rather than by decisions on the part of the DI recipient.

infer the demographic characteristics of compliers by exploiting regression discontinuities in demographic characteristics within the subsample of individuals who experienced a drop in replacement rates. We estimate that 13 percent of the sample consists of compliers. We find that, among compliers, the fraction who are female is 13 percentage points lower than the entire sample, the fraction who are fully disabled is 19 percentage points lower, and the fraction who are married is 13 percentage points higher (full results are in online Appendix Table A2). All of these differences are statistically significant. Thus, male, partially disabled, and married individuals were more likely to experience a reduction in their benefits than the sample as a whole. In line with these findings, we later split estimates out by gender, by degree of disability, and by marital status (in Tables 6 and A7).

Because we do not have income measures from before the reform, we cannot estimate the effects of the reform on income from labor and from other social assistance programs separately for those staying on DI and for those leaving DI.<sup>18</sup> However, we do have the replacement rate prior to the reexamination, which allows us to examine heterogeneity in the effects of the reform on replacement rates.

The first column of Table 3 shows the change in replacement rates at age 45.0 for those subject to the less stringent reexamination.<sup>19</sup> About 72 percent of this group saw no change in their replacement rate, 12 percent experienced an increase in their replacement rate, and 16 percent faced a decrease in their replacement rate. The second column shows the change in replacement rates at age 45.0 for those who underwent the more stringent reexamination. A much larger fraction (29 percent) in this latter group experienced a reduction in their replacement rate, and a much lower fraction (5 percent) saw an increase in their replacement rate. Still, even in the group subject to the more stringent reexamination, about two-thirds experienced no change in their replacement rate. The third column shows the treatment effect of the more stringent reexamination on the change in replacement rate, which is simply the difference between the first two columns. This column shows a downward shift in probability mass throughout the distribution of changes in replacement rates, showing that the reexamination made DI less generous throughout the distribution. Using the estimates of the causal effects of the more stringent reexamination on the full transition matrix of replacement rates (presented in online Appendix Table A3), we find that 70 percent of the reductions in replacement rates caused by the more stringent reexamination occurred among people who remained eligible for DI and that

<sup>18</sup> We do have covered earnings (but not income) prior to the start of the DI spell. Even though these earnings are, on average, more than a decade old and capped, we will later use them to get a rough sense of the heterogeneity of the impact of the more stringent reform on living standards.

<sup>19</sup> We run our standard reduced-form RD specification using dummies for each possible change in the replacement rate as the dependent variables. We estimate changes in the replacement rate at age 45.0 for those subject to the less stringent reexamination as the intercept of the right segment of the regression line at age 45.0. Changes in replacement rate for those subject to the more stringent reexamination are estimated by the intercept of the left segment of the regression line at age 45.0. More detailed estimates are provided in online Appendix Table A3, which presents the joint distribution of the replacement rate in 1996 (pre-reform) and 1999 (post-reform) at age 45.0 for those who were subject to the less stringent reexamination rules, as well as the impact of the reform on these joint probabilities (estimated by the RD effect). For example, the table shows that the more stringent reexaminations raised the fraction of fully disabled removed from the program from 9.2 percent ( $= 6.14/66.93$ ) to 11.8 percent ( $= (6.14 + 1.75)/66.93$ ). The estimates in Table 3 summarize all joint probabilities of online Appendix Table A3 by presenting the sums of the diagonal entries of the Appendix table.

TABLE 3—REEXAMINATIONS AND THE CHANGE IN THE REPLACEMENT RATE BETWEEN 1996 AND 1999

Change in the replacement rate	Predicted probability at age 45.0 for the less stringent reexamination (1)	Predicted probability at age 45.0 for the more stringent reexamination (2)	Treatment effect of the more stringent examination on the probability of the specified change in replacement rate (3)
7 "steps" less generous	6.14	7.89	1.75 (0.35)
6 "steps" less generous	0.46	1.56	1.09 (0.10)
5 "steps" less generous	0.85	2.41	1.56 (0.13)
4 "steps" less generous	1.08	2.25	1.17 (0.14)
3 "steps" less generous	1.09	2.30	1.21 (0.16)
2 "steps" less generous	2.17	3.69	1.52 (0.21)
1 "step" less generous	3.94	8.99	5.06 (0.29)
Same generosity	72.17	65.52	-6.65 (0.59)
1 "step" more generous	3.52	1.40	-2.12 (0.18)
2 "steps" more generous	1.46	0.67	-0.79 (0.12)
3 "steps" more generous	1.57	0.90	-0.66 (0.14)
4 "steps" more generous	1.87	0.91	-0.96 (0.14)
5 "steps" more generous	2.13	0.85	-1.28 (0.15)
6 "steps" more generous	1.54	0.66	-0.89 (0.12)

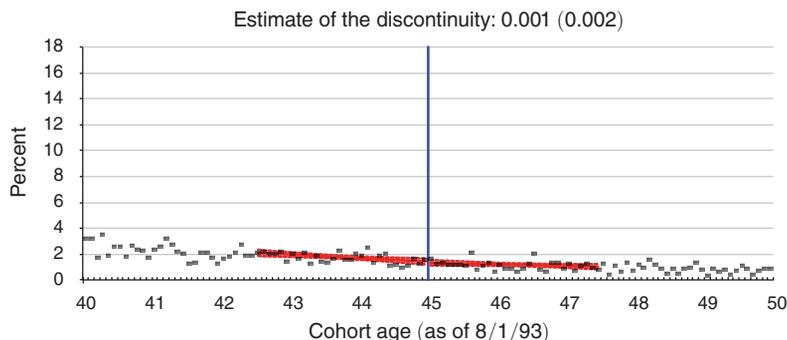
*Notes:* Standard errors are in parentheses. Each row is estimated using our standard reduced-form RD regression without demographic controls, where the outcome variable is a dummy for the change in the replacement rate between 1996 and 1999 that corresponds to row header. There are eight possible replacement rates (in percentages): 0, 14, 21, 28, 35, 42, 50.75, and 70, where we assign 0 to those who exit from DI before the postexamination replacement rate is recorded. The number of "steps" counts by how many distinct levels the replacement rate changed. For example, going to the next higher replacement rate (e.g., from 28 percent to 35 percent) is called a "1 step" increase in the generosity of the replacement rate. Column 1 shows the intercept at age 45.0 from the regression line to the right of the discontinuity (i.e., for those who underwent the less stringent reexamination), column 2 shows the intercept at age 45.0 from the regression line to the left of the discontinuity (i.e., for those who underwent the more stringent reexamination), and column 3 shows the treatment effect (i.e., the difference between columns 1 and 2). N=84,185. Age is measured as of August 1, 1993, so age is a measure for cohort.

the remaining 30 percent occurred among those who lost eligibility. These figures remain similar if we weight the causal effects on the reductions by the size of the reduction rather than by the number of people experiencing them.

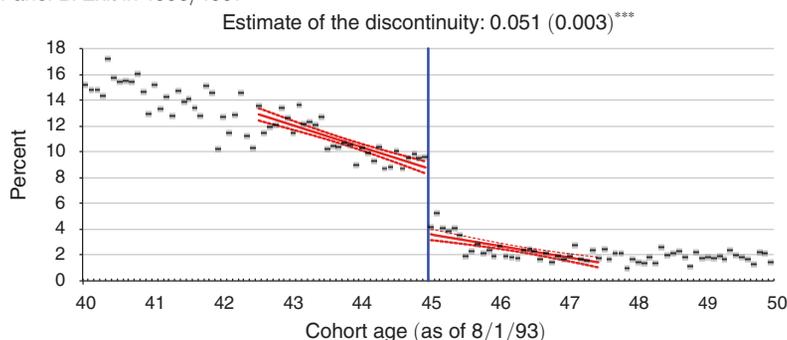
As an additional partial check on our identifying assumption that no factors besides the DI reexaminations had a discontinuous impact at age 45, Figure 2 reports DI exit rates separately for 1995, 1996/1997, 1998, and 1999. Exit is defined as the end date of the original DI spell, as recorded in the administrative data file, occurring during the year in question. We calculate these rates as fractions of DI claimants in our sample on January 1, 1996.<sup>20</sup> Since no reexaminations took place in 1995 for DI claimants in the age 40+ cohorts, a discontinuity at age 45 in the 1995 exit rate would invalidate our identifying assumption. Reassuringly, the 1995 exit rate shows no sign of a discontinuity at age 45. In 1996 and 1997, the age 40–44 cohort was reexamined as well as part of the age 45 cohort. In precisely these years, the discontinuity at age 45 is very pronounced. In 1998, the remainder of the age 45 cohort and some of the age 46 cohort were reexamined, which explains the statistically significant discontinuity in the opposite direction. This discontinuity, however, is much smaller in size than those in other years because the age 45+ cohort was reexamined under the older and less stringent standards. Hence, if we calculate the

<sup>20</sup>However, the exit rate for 1995 is calculated as a fraction of DI recipients as of January 1, 1995.

Panel A. Exit in 1995



Panel B. Exit in 1996/1997

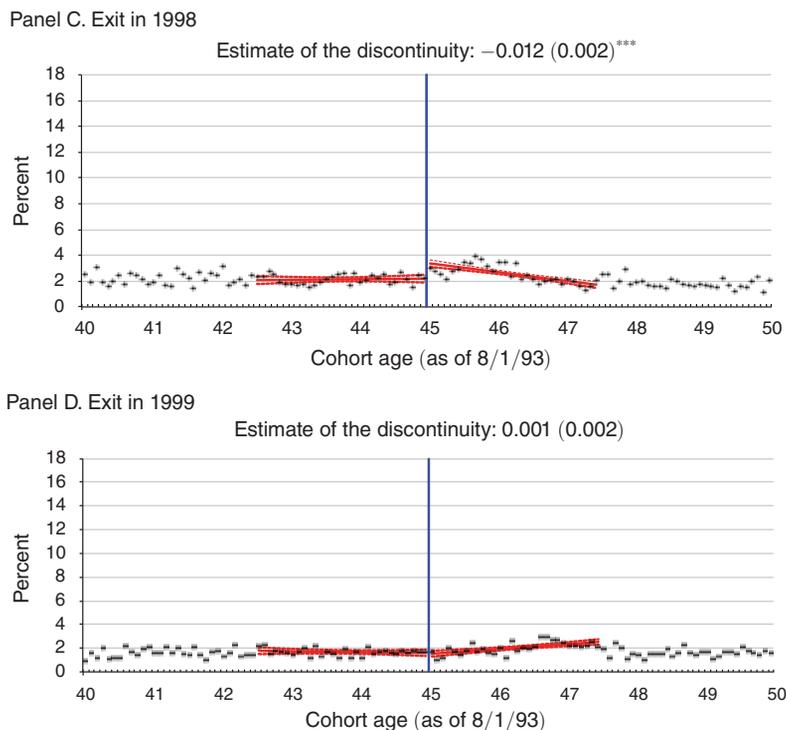
FIGURE 2. EXIT FROM DI BY YEAR (*Continued*)

total exit rate over the 1996–1998 period, we find a discontinuous increase in exit for the group subject to the more stringent reexaminations. In 1999, all the reexaminations for the age 44–45 cohort were completed, and we find no discontinuity in exit rates at the age cutoff.

### B. Reduced-Form Impacts on Labor Market and Social Assistance Outcomes

To what extent did individuals whose DI was reduced by the reform end up in other social assistance programs, and to what extent did they find paid work? The answer to this question is critical for judging the effectiveness of the reform. In the former case, the reform merely shuffles individuals across programs, and budgetary savings only occur to the extent that benefits in other programs are lower than DI benefits. In the latter case, increased earnings are an indication of moral hazard among existing disability recipients. In this subsection, we examine labor market and social assistance outcomes in 1999 (about two years after the reexaminations took place), which is the first year for which we have the required data. In subsection IIIF, we will examine the effects over a longer horizon.

We start by analyzing the reduced-form effects of the more stringent reexamination (and the associated benefit cuts) on receipt of other forms of social assistance. The first panel of Figure 3 plots income from social assistance other than DI by cohort. The figure shows an upward jump in income from other social assistance

FIGURE 2. EXIT FROM DI BY YEAR (*Continued*)

Notes: Standard errors are in parentheses. The dotted lines represent the 95 percent confidence intervals. The exit rate is defined as a fraction of our sample in January 1996, except for panel A where it is a fraction of the sample in January 1995. Regression estimates come from reduced-form RD regressions without demographic control variables.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

programs for the cohort that underwent the more stringent reexamination. In fact, the RD regression estimates that the more stringent reexamination increased other social assistance income by €305 per year. The second panel shows that the fraction receiving any social assistance income from a source other than DI discontinuously increases by 4.6 percentage points at the age cutoff for the more stringent reexaminations. Both increases are highly significant and represent an increase of about one-third in relative terms. In other words, we find clear evidence of substitution of other forms of social assistance for DI benefits.

Do people who leave DI fully account for the increased income from other social assistance programs, or do those who remain on DI also receive more income from other social assistance programs? To answer this question, we first note that there are two types of individuals who exited from DI: (i) those who exited from DI under the more stringent reexamination but would not have exited from DI under the less stringent reexamination (*marginal leavers*) and (ii) those who would have exited anyway (*inframarginal leavers*). Panel C of Figure 1 allows us to determine the population fractions of marginal and inframarginal leavers. Just to the right of the

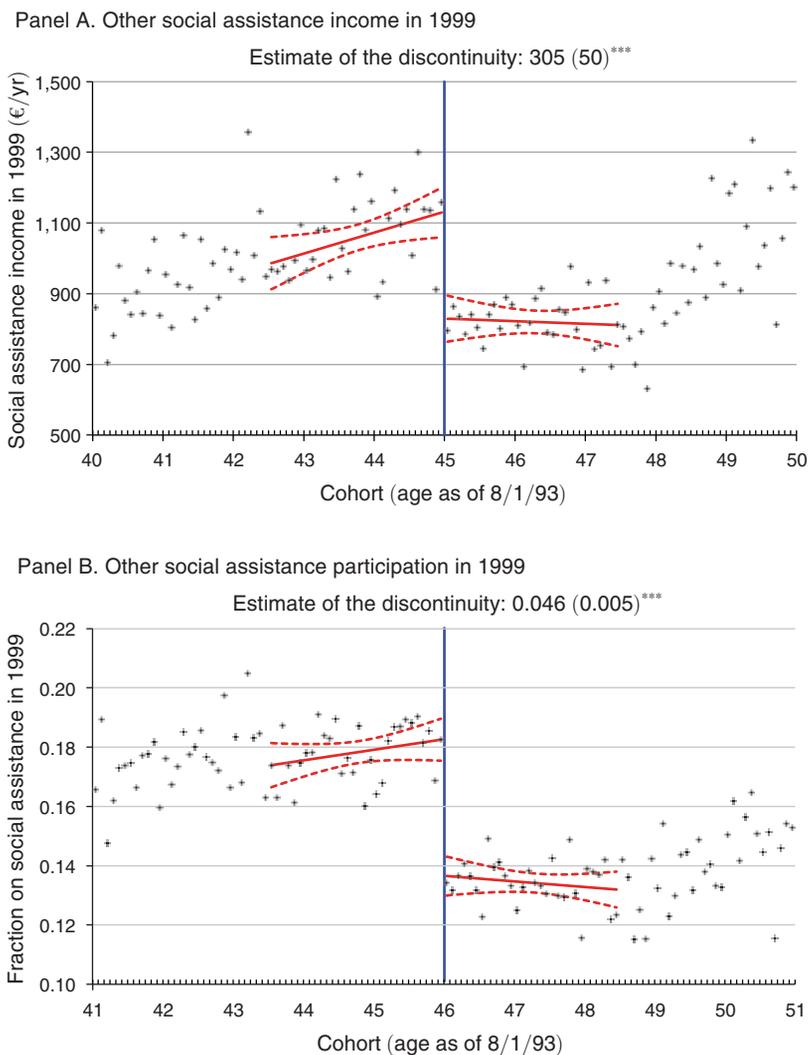


FIGURE 3. EFFECTS OF DI REFORM ON SOCIAL ASSISTANCE OTHER THAN DISABILITY INSURANCE

*Notes:* Standard errors are in parentheses. Figures are based on 84,185 observations. The dotted lines represent the 95 percent confidence intervals. Social Assistance Income excludes any benefits from DI (both from the original spell or any new spell). Similarly, Social Assistance Participation excludes participation in DI. Regression estimates come from regressions without demographic control variables.

<sup>\*\*\*</sup> Significant at the 1 percent level.

<sup>\*\*</sup> Significant at the 5 percent level.

<sup>\*</sup> Significant at the 10 percent level.

age discontinuity, individuals faced the less stringent reexamination and all leavers are therefore inframarginal leavers by definition. The DI participation rate just to the right of the age continuity is 93.2 percent (see panel C of Figure 1), so 6.8 percent ( $= 100 - 93.2$ ) of individuals are inframarginal leavers. Individuals just to the left of the age discontinuity faced the more stringent reexamination, so leavers just to the left consist of both inframarginal and marginal leavers. The participation rate just to the left of the discontinuity is 89.4 percent, which implies that inframarginal

and marginal leavers combined comprise 10.6 percent ( $= 100 - 89.4$ ) of the population. Hence, marginal leavers make up 3.8 percent ( $= 10.6 - 6.8$ ). This means that 35.8 percent ( $= 3.8/10.6$ ) of all leavers to the left of the discontinuity are marginal leavers and 64.2 percent are inframarginal leavers.

We note that inframarginal leavers cannot be responsible for the jump in social assistance receipt observed in panel A of Figure 3, because inframarginal leavers, by definition, are not affected by an increase in the stringency of the reexamination (they would have exited in any case). So, to the extent that this jump is caused by leavers, it must be caused by marginal leavers. To find the degree to which this jump can be attributed to marginal leavers, we rerun the RD regression of panel A of Figure 3, but now on the subsample of those who left DI. The resulting regression line (not shown) just to the right of the age cutoff lies at €1,962 of other social assistance income per year. Because just to the right of the age cutoff all leavers are inframarginal by definition, this implies that other social assistance income is equal to €1,962 per year for inframarginal leavers. Just to the left of the age cutoff, where we have both marginal and inframarginal leavers, the regression line lies at €2,564 per year. Thus, the weighted average of income from other social assistance programs for all leavers to the left of the discontinuity must equal €2,564, which implies that income from other social assistance for marginal leavers must have been €3,644 per year (so that 35.8 percent of €3,644 plus 64.2 percent of €1,962 equals €2,564). If marginal leavers receive €3,644 per year from other social assistance, the maximum amount by which their income from other social assistance programs could have possibly increased due to the more stringent reexamination is €3,644. Given that marginal leavers comprise only 3.8 percent of the entire sample (so both stayers and leavers) at the age cutoff, they can, at most, account for an increase of  $0.038 \times 3,644 = €138$  in the RD estimate of €305/year that we found for the entire sample. Thus, we conclude that those leaving DI altogether can be responsible for at most 43 percent ( $= 138/305$ ) of the overall jump in income from other social assistance programs, and that at least 57 percent must be due to increases in other social assistance income among those remaining on DI.

In Figure 4, we examine the extent to which individuals subject to the stricter reexaminations were more likely to receive income from new DI spells. In 1999, this occurred only to a minimal extent. Among those facing the more stringent reexamination, annual income from new DI spells is a statistically insignificant €8 higher. The participation rate in new DI spells is 0.2 percentage points higher, which is statistically significant but still small relative to the 3.8 percent of DI recipients who were induced to leave due to the more stringent reexamination. As we will show below, reentry into DI becomes more important over time.

Next, we present the reduced-form effects of the more stringent reexamination on labor market outcomes. The first panel of Figure 5 plots earnings (including self-employment income) in 1999 by cohort. The figure shows a discontinuity in earnings at the cutoff age, but the discontinuity is not as visually compelling as in the earlier figures. The RD regression estimates that earnings are €624 per year higher at the cutoff age for those who were subject to the more stringent reexamination, and this estimate is highly significant. The €624 increase represents an

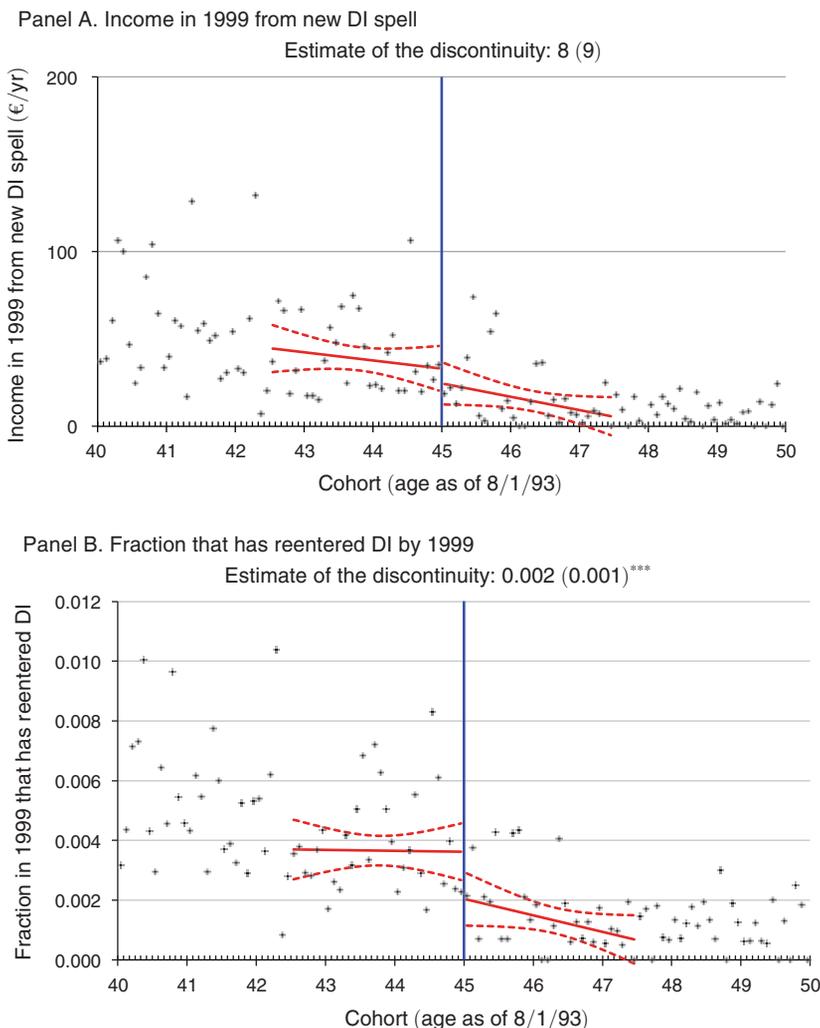


FIGURE 4. EFFECTS OF DI REFORM ON REENTRY INTO DI

Notes: Standard errors are in parentheses. Figures are based on 84,185 observations. The dotted lines represent the 95-percent confidence intervals. Only DI income and participation in DI that originate from spells that started after the reexamination are included in this figure. Regression estimates come from regressions without demographic control variables.

- \*\*\* Significant at the 1 percent level.
- \*\* Significant at the 5 percent level.
- \* Significant at the 10 percent level.

11 percent increase in annual earnings. This figure establishes our qualitative finding that disability income crowds out labor income. It also contributes to the literature on the labor supply disincentive effects of disability insurance by showing clear evidence of labor supply responses among prime-age DI recipients who are long-term recipients of DI (durations of at least two years at the time of the reform, but of ten years on average). We will discuss the economic magnitude of the labor supply response in the next subsection.

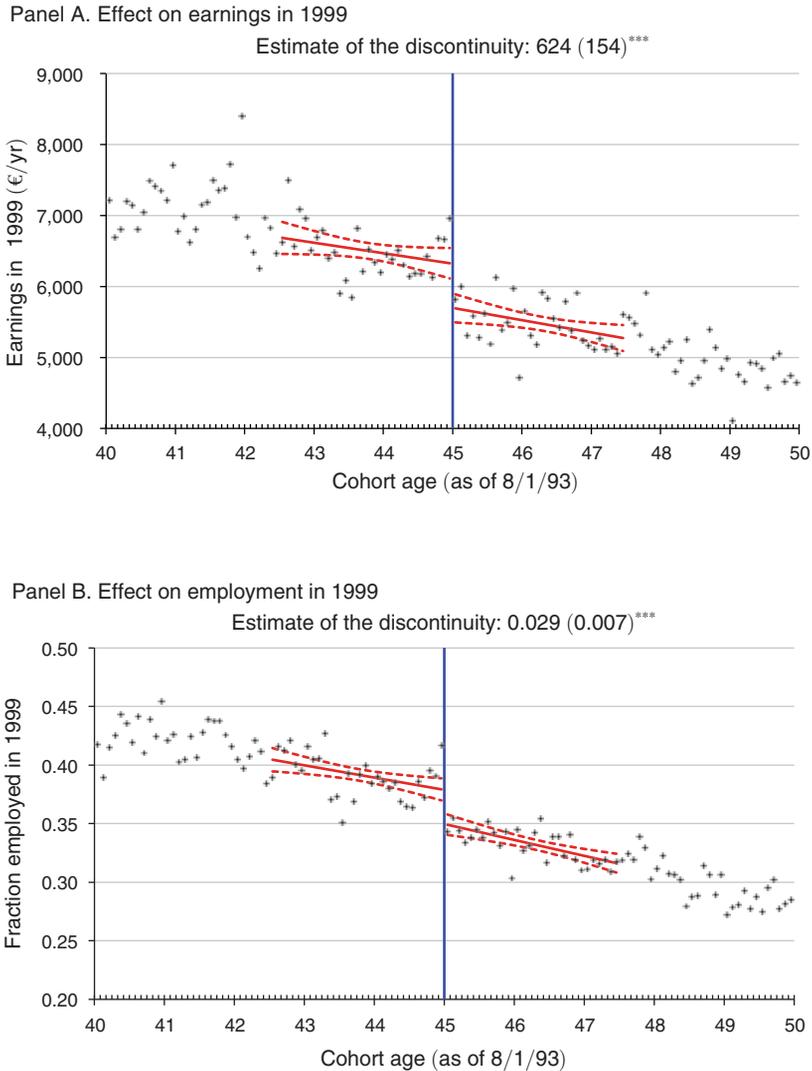


FIGURE 5. EFFECTS OF DI REFORM ON LABOR MARKET OUTCOMES

*Notes:* Standard errors are in parentheses. Figures are based on 84,185 observations. The dotted lines represent the 95 percent confidence intervals. Employment is defined as having positive earnings from employment or self-employment. Regression estimates come from reduced-form RD regressions without demographic control variables.

<sup>\*\*\*</sup> Significant at the 1 percent level.

<sup>\*\*</sup> Significant at the 5 percent level.

<sup>\*</sup> Significant at the 10 percent level.

Because we do not have earnings for prior years, we cannot precisely determine the extent to which the average increase in earnings stems from nonworkers finding employment (extensive margin) or from workers increasing their earnings (intensive margin). However, at least some of the increase comes from the extensive margin because the second panel of Figure 5 shows a clear discontinuity in the fraction of individuals with strictly positive income from wages or self-employment.

The RD regression estimates that the more stringent reexamination caused the fraction employed to increase by 2.9 percentage points. To explain the observed increase in earnings in the absence of an intensive-margin labor supply response, average earnings for those who started working would need to be €21,500 ( $= 624/0.029$ ) per year, which is higher than the observed average earnings for those with positive earnings (€17,000 per year). It therefore seems likely that some of the response also occurred along the intensive margin. When we do a bounding calculation similar to the one we did above for income from other social benefits, we find that those who left DI altogether can be responsible for at most 61 percent of the overall jump in earnings caused by the more stringent reexamination, and that at least 39 percent must be due to earnings increases among those remaining on DI.

### C. Benefit Substitution and Earnings Crowd-Out

Figures 3 and 5 establish that people substitute between DI income and other forms of social assistance, and that DI benefits crowd out labor income. We now turn to the economic magnitudes of earnings crowd-out and substitution of social assistance. In the first column of Table 4, we scale our reduced-form estimates by the amount by which disability benefits from the original spell decrease at the age discontinuity, whereas in the second column, we scale by the discontinuity in the replacement rate. We implement this scaling by running IV regressions following the standard “fuzzy” RD specification.<sup>21</sup> We include a rich set of demographic control variables to increase the precision of the estimates. As should be the case with a valid RD design, the control variables do not substantially affect the magnitudes of our estimates (see online Appendix Table A4 for the corresponding estimates without controls). Given that the reexamination was, in all respects, more stringent for those below the cutoff age, the monotonicity assumption required for the fuzzy RD design should be satisfied; being subject to a reexamination following the more stringent new protocol rather than the old protocol weakly decreases the benefit amount and weakly decreases the replacement rate at any level of earnings for any given individual (despite the increase in the earnings exemption). We do not interpret these IV estimates as causal impacts of the level of DI benefits per se or as causal estimates of the DI replacement rate per se because the reduction in benefits and replacement rates was not uniform. In fact, the reduction varied across individuals and by earnings level for any given individual. The response to the benefit cuts is likely not only a function of the level of the cuts, but also of the heterogeneity in the cuts. We therefore see the IV estimates as a way to relate the magnitudes of the behavior effects of the reform to alternative measures of the size of the reform. In other words, we view this primarily as a scaling exercise.

Panel A of Table 4 shows that there is only a minimal increase in income from or participation in new DI spells in the short term (roughly two years after the reexaminations) in response to the benefit and replacement rate reductions. For example, for every €1,000 reduction in existing DI benefits, the increased expenditure on new

<sup>21</sup> Excellent discussions of the theory and practice of RD methods can be found in Hahn, Todd, and van der Klaauw (2001), Imbens and Lemieux (2008), and Lee and Lemieux (2010).

TABLE 4—EARNINGS CROWD-OUT AND BENEFIT SHIFTING

	Effect scaled by decrease in amount of original DI (in 1,000 €/year)	Effect scaled by decrease in the replacement rate (fraction)
<i>Panel A. New DI benefit in 1999</i>		
Income from a new DI spell	0.008 (0.008)	0.142 (0.148)
Participation dummy	0.001 (0.001)**	0.026 (0.011)**
<i>Panel B. Other social assistance in 1999</i>		
Income from other social assistance	0.297 (0.047)***	5.212 (0.787)***
Participation dummy	0.044 (0.005)***	0.778 (0.081)***
<i>Panel C. Labor market outcomes in 1999</i>		
Earnings	0.618 (0.108)***	10.848 (1.924)***
Employment dummy	0.029 (0.005)***	0.511 (0.084)***
<i>Panel D. Total</i>		
Income except from original DI spell	0.923 (0.113)***	16.201 (1.983)***
Dummy for work or other social assistance	0.057 (0.006)***	0.992 (0.092)***

*Notes:* Standard errors are in parentheses. Each entry in the table comes from a separate IV regression based on the fuzzy RD design. The dependent variable is listed in the rows. Other social assistance income excludes any benefits from DI (both from the original spell or any new spell). The variable that is instrumented (endogenous explanatory variable) is listed in the columns. The instrument itself is the treatment dummy (age as of August 1, 1993 less than 45). Earnings and income are measured in thousands of euros per year. The replacement rate is expressed as a fraction. Each regression is based on 84,185 observations. The following controls are used in the regressions: age in months as of August 1, 1993, [age–45] interacted with the treatment dummy, six dummies for degree of disability in 1996, a cubic polynomial in pre-DI earnings, nine national origin dummies, a dummy for being married in 1996, 39 regional dummies, a cubic polynomial in duration in DI at the start of the reform, a full set of interactions between the dummies for the degree of disability and the cubic polynomial in pre-DI earnings, a gender dummy, and a full set of interactions between all previously listed controls and gender. In total, each regression has 163 control variables.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

DI spells is €8 (not significant). Panel B examines the extent to which the reduced generosity of disability benefits induces individuals to shift to other forms of social assistance. The first row in column 1 of panel B shows that, for a €1 reduction in disability benefits, individuals receive €0.30 more from other social assistance programs in 1999. Thus, the benefit-substitution ratio is 0.30. A government not taking this substitution into account would overestimate the reduction in government expenditure from tightening the DI eligibility rules by 43 percent. The second row shows that for each €1,000 decrease in annual DI benefits caused by the more stringent rules, the probability that an individual receives income from another social assistance program increases by 4.4 percentage points. An alternative way of scaling the degree of substitution between social assistance programs is provided in the second column, which shows that, for a 10-percentage-point reduction in disability replacement rates, income from other social assistance programs increases by €521 per year (an increase of more than 50 percent) and the probability of participation in

other social assistance programs increases by 7.8 percentage points. The estimates of panel B in Table 4 establish that benefit substitution is not only statistically significant, but also important in economic terms.

Benefit substitution can occur mechanically when individuals automatically receive more income from other social assistance programs as their DI benefits decrease. While this might explain some of the substitution, it cannot account for the entire reaction, because we also observe individuals enrolling in other forms of social assistance for which enrollment is not automatic. Benefit substitution can also be a result of the individual actively searching for alternative sources of benefits and trying to qualify for them. Finally, benefit substitution can occur when caseworkers steer individuals toward alternative sources of support. We have no direct evidence on the relative importance of these three channels, but we suspect that all three may have contributed to the observed amount of benefit substitution to some degree.

The estimate in the first row in column 1 of panel C in Table 4 indicates that, for each euro of benefits decrease caused by the more stringent reexamination, the more stringent reexamination induced individuals to increase earnings by €0.62 in 1999. In other words, we find an earnings crowd-out ratio of 0.62: one euro of DI benefits crowds out €0.62 of earnings.<sup>22</sup> Alternatively, one can scale the change in earnings by the change in total benefits (including DI benefits) due to the more stringent reexamination. For each euro in decreased DI benefits on the original spell, other benefits went up by €0.30 and DI benefits on new spells went up by €0.01. Therefore, total benefits only decreased by €0.69. Thus, per euro decrease in total benefits, earnings went up by €0.90 ( $= 0.62/0.69$ ). The second row of panel C examines the extensive margin response and shows that, per €1,000 of disability benefits decrease caused by the more stringent reexamination, the probability of being employed in 1999 increases by 2.9 percentage points. The second column presents analogous estimates, but now scaled by the change in replacement rates caused by the more stringent reexamination. We find that, for a ten-percentage-point decrease in replacement rates, earnings increase by €1,085 per year (or about 19 percent) and the probability of employment increases by 5.1 percentage points. All four estimates in panel C are highly statistically significant and establish that the degree to which DI benefits crowd out labor market earnings and participation is economically meaningful.

In the late 1990s, the Netherlands experienced an economic boom, which likely made it relatively easy for individuals to increase their labor supply and reduced incentives to look for other forms of social assistance. As a result, our estimate for earnings crowd-out may be higher than it would have been during average economic times, while the estimate of benefit substitution may be lower. During the early 2000s, the Netherlands experienced a recession, and, as we will see in Section III F below, our estimates of earnings crowd-out and benefit substitution remained similar during that period. A further reason why the earnings crowd-out estimate may be relatively high is that the Netherlands had various policies to foster reintegration into the labor market.

<sup>22</sup> All amounts are gross of tax and social insurance contributions. Both benefits and earnings are subject to taxation and mandatory social insurance contributions, so we can directly compare changes in income to changes in benefits.

DI recipients have an earnings exemption which equals their indexed previous earnings times the degree to which they were deemed able to work (i.e., one minus the degree of disability). Any earnings beyond the exemption faced a high implicit tax rate. Thus, if the reexamination led to a reduction in the disability rating, this both reduced the DI benefit (an income effect) and increased the earnings exemption (a substitution effect).<sup>23</sup> Therefore, like most of the previous literature on the labor supply response to DI, we cannot determine the extent to which the response is driven by the substitution effect or by the income effect.<sup>24</sup> However, given the large magnitude of the earnings reaction (especially if compared to the change in total benefits), we suspect that incentive effects stemming from the change in the earnings exemption played a large role.

Panel D of Table 4 presents the combined effect of benefit substitution and earnings crowd-out. The estimate in the first row and column indicates that individuals increased income from other social assistance (including new DI spells) and work by €0.92 per euro of original DI benefits lost. In other words, on average, individuals almost fully offset the decrease in DI benefits by increasing income from other sources, and we cannot reject the hypothesis that the lost DI was completely offset ( $p$ -value 0.494). Even if the offset is complete, individuals now have to work more and are worse off to the extent they receive disutility from supplying labor.<sup>25</sup> The second row shows the effect on a dummy for working or receiving income from a social assistance program other than the original DI spell. We find that per €1,000 decrease in DI, an individual is 5.7 percentage points more likely to obtain income from a new source. The fact that this estimate is less than the sum of the estimates in row 2 of panels A, B, and C indicates that some individuals both started working and started drawing income from other forms of social assistance. In particular, per €1,000 decrease in DI, individuals became 1.7 ( $= 0.1 + 4.4 + 2.9 - 5.7$ ) percentage points more likely to have both income from other social assistance programs and labor income in 1999.

The estimates in Table 4 are based on a piece-wise linear specification with a bandwidth of  $\pm 2.5$  years around the cutoff age, which is the bandwidth suggested by applying the Imbens-Kalyanaraman criterion (2012) to our data. Online Appendix Table A6 explores the sensitivity of our key results to the choice

<sup>23</sup>Note, however, that there is no level of earnings at which DI benefits are increased by a reduction in the disability rating, whereas there is always some level of earnings at which DI benefits are strictly decreased by a reduction in the disability rating. Thus, despite the increase in the earnings exemption, a reduction in the disability rating implies a reduction in DI generosity.

<sup>24</sup>Important exceptions are Campolieti and Riddell (2012), who show a labor supply response to the increase in the earnings exemption (a pure substitution effect), and Marie and Vall Castello (2012), who measure a labor supply response to a pure increase in DI benefits (an income effect).

<sup>25</sup>Even if we cannot reject that individuals are, on average, able to fully offset the DI benefit cut, it is conceivable that certain subgroups are not able to do so. To examine this possibility, online Appendix Table A5 compares current income to earnings (adjusted for general wage growth) prior to the start of the DI spell. To briefly summarize our findings: we find that, on average, 65.9 percent of the sample had a lower income in 1999 than their earnings prior to DI. The RD estimate indicates that this fraction is a statistically significant 1.4 percentage points lower for those subject to the more stringent reexamination. Therefore, the more stringent reexamination slightly increased the fraction experiencing an income increase. On the other hand, 9.9 percent of the sample had an income in 1999 that was less than 50 percent of earnings prior to DI. This percentage is a statistically significant 1.6 percentage points higher for those subject to the more stringent reexamination. Hence, the effect of the more stringent reexamination on income changes is heterogeneous. It raises both the fraction experiencing increases in income and the fraction experiencing sharp drops in income.

of bandwidth and to the choice of polynomial in the running variable. Overall, the table indicates that our qualitative results are reasonably robust to the choice of bandwidth and functional form for the running variable, though the point estimates do vary a fair amount across specifications.

#### *D. Effects by Income Source, Gender, and Degree of Disability*

The first column of panel A of Table 5 splits out the results of the first column of Table 4 by source of income. Of the increase in benefits, 67 percent came from increased UI benefits, 10 percent from increased General Assistance, 3 percent from reentry into DI, and 20 percent from all other types of social support benefits. The large increase in UI benefits was not automatic, but individuals did build up “work history” for the purpose of UI benefits while receiving DI (even if not working). We find that increased wage earnings account for 80 percent ( $= 0.492/0.618$ ) of the earnings response, while changes in self-employment income account for 20 percent. Given that evasion of wage earnings is hard and limited, this breakdown indicates that the earnings response is unlikely to be largely driven by people not changing their actual labor supply but simply starting to report their earnings. Panel B examines receipt of any amount by income sources, and its results are very similar to the results of panel A.

Columns 2 and 3 of Table 5 split out the results of the first column by gender. These columns suggest that social support substitution is more predominant among women. In particular, the point estimate of the benefit-substitution ratio is much larger for women than for men (0.48 versus 0.26), but this difference is not statistically significant ( $p$ -value 0.149). However, the difference is statistically significant if we look at the participation response for other forms of social assistance. Per €1,000 decrease in DI benefits, women increase their participation in other social assistance programs by 7.4 percentage points, which is nearly twice the 3.8-percentage-point increase by men. In response to a given DI benefit cut, women are also significantly more likely than men to start working, whereas the point estimate of the earnings response is actually slightly larger for men than for women (but this difference is not statistically significant). The fact that labor force participation is only 18 percent for women but 45 percent for men may explain why women experience a larger response on the extensive margin while total earnings increase slightly more for men because the scope for an intensive-margin response is larger among men. There is no significant difference in the degree to which men and women are able to offset the decrease in DI benefits by other sources of income. We cannot determine what exactly drives differences in the effects on men and women, but we suspect that likely explanations are differences in initial DI benefit levels, differences in types of disabilities, and differences in opportunities in the labor market and in household production.

Table 6 analyzes benefit substitution and earnings crowd-out by the degree of disability into which individuals were classified as of January 1, 1996 (before the reexaminations took place).<sup>26</sup> Panel A of Table 6 shows that the effects on benefits

<sup>26</sup>We also investigated whether the benefit-substitution ratio and the labor crowd-out ratio varied by marital status, previous earnings, duration of the DI spell, and national origin. We found no significant differences along

from new DI spells are small and do not vary significantly by gender or disability rating. Panel B shows that social support substitution is much more prevalent among individuals classified as fully disabled than among those classified as partially disabled. This finding applies both for social assistance benefit amounts and for social assistance participation, and it holds in the entire sample as well as the subsamples by gender. These differences are not only statistically significant but also large in magnitude. The benefit-substitution ratio for the fully disabled is 0.49, which is four times as large as the ratio of 0.12 for partially disabled recipients. About 90 percent of this difference in benefit substitution ratios is accounted for by the larger increases in UI benefits among the fully disabled relative to partially disabled individuals.

In contrast, we find high rates of earnings crowd-out both for those classified as partially disabled and for those classified as fully disabled. Panel C shows that the point estimate of crowd-out is somewhat higher for the partially disabled than for the fully disabled (0.68 versus 0.52), but this difference is not statistically significant. Perhaps the high degree to which the fully disabled are able to replace foregone disability income with earnings would not be surprising if the cut in disability income were minimal. This, however, is not the case. The more stringent reexamination caused a €799 (standard error: 86) reduction in benefits among the fully disabled. While this figure is only about half of the corresponding figure for the partially disabled (€1,678, standard error: 126), it is still a meaningful reduction in benefits. Moreover, about 30 percent of the fully disabled who experienced a benefit cut because of the more stringent reexamination became ineligible for DI. In short, given that the fully disabled experienced meaningful cuts in benefits, the high degree to which they were able to replace foregone disability income with labor income is quite striking. However, it should be kept in mind that the disability rating depended in part on the existence of sufficient suitable occupations in the applicant's region, and that an applicant could be classified as fully disabled if this criterion was not met.

Panel D of Table 6 shows that, on average, both the partially and fully disabled are able to offset basically all of their lost DI income by other sources of income. The point estimates indicate that the fully disabled actually offset somewhat more of the lost DI benefits than the partially disabled, but this difference is not statistically significant.

---

these dimensions when we cut each dimension as closely as possible into two equally sized groups. See online Appendix Table A7 for details. If we cut duration into three groups of roughly equal size (longer than ten years, between four and ten years, and less than four years as of August 1, 1993), then we find that the labor supply response among those with durations shorter than four years is significantly larger than the labor supply response among those with durations between four and ten years. However, we continue to find a significant labor crowd-out ratio even for those who have been on DI for over ten years. We cannot stratify our estimates by medical diagnosis because much of this data was added retroactively and, consequently, is significantly less likely to be missing for those who remained on DI. However, we can in principle use those remaining on DI to infer differential exit by medical diagnosis. The medical part of the reexamination was the same on either side of the age cutoff, which implies there should be no causal effect of the more stringent reexamination on the medical diagnosis itself. Therefore, any discontinuity at the age cutoff in the prevalence of a given medical diagnosis among those who remained on DI must be due to differential exit by that medical diagnosis. Unfortunately, we lack statistical precision on this inferred distribution of differential exit by medical diagnosis, as detailed in online Appendix Table A8.

TABLE 5—EARNINGS CROWD-OUT AND BENEFIT SHIFTING BY SOURCE AND BY GENDER

	Effect of reform per 1,000 €/year decrease in amount of original DI			<i>p</i> -value gender dif.
	Full sample	Males	Females	
<i>Panel A. Amounts in 1999</i>				
Total other social assistance, of which:	0.305 (0.047)***	0.261 (0.047)***	0.482 (0.146)***	0.149
Unemployment insurance	0.203 (0.020)***	0.187 (0.021)***	0.266 (0.051)***	
General assistance	0.030 (0.007)***	0.032 (0.007)***	0.024 (0.024)	
Reentry into DI	0.008 (0.008)	−0.001 (0.009)	0.046 (0.025)*	
All other benefits	0.063 (0.041)	0.043 (0.040)	0.146 (0.129)	
Total earnings, of which:	0.618 (0.108)***	0.632 (0.124)***	0.564 (0.208)***	0.781
Wage earnings	0.492 (0.098)***	0.497 (0.114)***	0.471 (0.181)***	
Self-employment earning	0.126 (0.061)**	0.135 (0.070)*	0.093 (0.113)	
Total income from other income sources	0.923 (0.113)***	0.892 (0.127)***	1.046 (0.250)***	0.582
<i>Panel B. Participation in 1999</i>				
Any income from other social assistance	0.045 (0.005)***	0.038 (0.005)***	0.074 (0.016)***	0.037
Any unemployment insurance	0.034 (0.003)***	0.030 (0.003)***	0.050 (0.009)***	
Any general assistance	0.009 (0.002)***	0.010 (0.002)***	0.007 (0.006)	
Reentry into DI	0.001 (0.001)**	0.001 (0.001)*	0.003 (0.002)	
Any other benefits	0.015 (0.004)***	0.010 (0.004)**	0.035 (0.013)***	
Any work	0.029 (0.005)***	0.023 (0.005)***	0.053 (0.013)***	0.026
Any wage income	0.024 (0.005)***	0.018 (0.005)***	0.050 (0.012)***	
Any self-employment income	0.006 (0.003)**	0.007 (0.003)**	−0.001 (0.005)	
Any other income source	0.057 (0.006)***	0.044 (0.006)***	0.105 (0.019)***	0.002

*Notes:* Standard errors are in parentheses. Each entry in the table comes from a separate IV regression based on the fuzzy RD design. The dependent variable is listed in the rows. Income and earnings are measured in thousands of euros per year. “Other income sources” excludes DI income from the original spell. General assistance provides an income floor for everyone and does not require having dependents. Any other benefits are benefits from a large number of smaller (about 30) benefit programs. The variable that is instrumented (endogenous explanatory variable) is the amount of DI, so all coefficients can be interpreted as effect size per €1,000/year decrease in DI. The instrument itself is the treatment dummy (age less than 45 as of August 1, 1993). The regressions are based on 84,185; 55,772; and 28,413 observations for the full sample, males, and females, respectively. See the note to Table 4 for the demographic controls included in the regression.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

TABLE 6—EARNINGS CROWD-OUT AND BENEFIT SHIFTING BY DEGREE OF DISABILITY AND GENDER

	Effect of reform per 1,000 €/year decrease in amount of original DI			<i>p</i> -value gender difference
	Full sample	Males	Females	
<i>Panel A. New DI benefit in 1999</i>				
Income from a new DI spell				
Partially disabled in 1996	0.005 (0.008)	0.001 (0.008)	0.021 (0.025)	0.457
Fully disabled in 1996	0.010 (0.015)	−0.005 (0.015)	0.063 (0.042)	0.124
<i>p</i> -value on difference by disability	0.770	0.694	0.388	
Participation dummy				
Partially disabled in 1996	0.001 (0.001)	0.001 (0.001)	0.001 (0.002)	0.971
Fully disabled in 1996	0.002 (0.001)*	0.001 (0.001)	0.005 (0.003)	0.235
<i>p</i> -value on difference by disability	0.419	0.954	0.283	
<i>Panel B. Other social assistance in 1999</i>				
Income from other social assistance				
Partially disabled in 1996	0.117 (0.047)**	0.123 (0.049)**	0.091 (0.141)	0.833
Fully disabled in 1996	0.491 (0.086)***	0.421 (0.084)***	0.740 (0.263)***	0.247
<i>p</i> -value on difference by disability	<0.001	0.002	0.029	
Participation dummy				
Partially disabled in 1996	0.015 (0.006)***	0.015 (0.006)***	0.012 (0.015)	0.853
Fully disabled in 1996	0.076 (0.010)***	0.061 (0.009)***	0.128 (0.035)***	0.064
<i>p</i> -value on difference by disability	<0.001	<0.001	0.002	
<i>Panel C. Labor market outcomes in 1999</i>				
Earnings				
Partially disabled in 1996	0.682 (0.166)***	0.732 (0.186)***	0.435 (0.357)	0.460
Fully disabled in 1996	0.520 (0.128)***	0.506 (0.148)***	0.572 (0.250)**	0.819
<i>p</i> -value on difference by disability	0.441	0.341	0.753	
Employment dummy				
Partially disabled in 1996	0.023 (0.006)***	0.020 (0.006)***	0.037 (0.017)**	0.357
Fully disabled in 1996	0.034 (0.008)***	0.026 (0.009)***	0.062 (0.019)***	0.087
<i>p</i> -value on difference by disability	0.246	0.555	0.333	

(Continued)

### E. Responses of Partners of DI Recipients and Total Budgetary Impacts

In Table 7 we provide estimates of benefit substitution and earnings crowd-out at the household level. These estimates differ from our baseline estimates of Table 4 in that the current estimates account for possible responses of partners of (former)

TABLE 6—EARNINGS CROWD-OUT AND BENEFIT SHIFTING BY DEGREE OF DISABILITY AND GENDER (*Continued*)

	Effect of reform per 1,000 €/year decrease in amount of original DI			<i>p</i> -value gender difference
	Full sample	Males	Females	
<i>Panel D. Total</i>				
Income except from original DI spell				
Partially disabled in 1996	0.804 (0.163)***	0.856 (0.182)***	0.547 (0.360)	0.444
Fully disabled in 1996	1.021 (0.153)***	0.921 (0.166)***	1.376 (0.392)***	0.285
<i>p</i> -value on difference by disability	0.332	0.793	0.119	
Dummy for work or other soc. asst.				
Partially disabled in 1996	0.030 (0.005)***	0.028 (0.005)***	0.036 (0.017)**	0.656
Fully disabled in 1996	0.084 (0.011)***	0.062 (0.011)***	0.163 (0.041)***	0.017
<i>p</i> -value on difference by disability	<0.001	0.005	0.004	
Observations:				
Partially disabled in 1996	28,509	22,590	5,919	
Fully disabled in 1996	55,676	33,182	22,494	

*Notes:* Standard errors are in parentheses. Each entry in the table comes from a separate IV regression based on the fuzzy RD design. The dependent variable is listed in the rows. Income and earnings are measured in thousands of euros per year. The variable that is instrumented (endogenous explanatory variable) is the amount of DI, so all coefficients can be interpreted as effect size per €1,000/year decrease in DI. The instrument itself is the treatment dummy (age less than 45 as of August 1, 1993). Degree of disability is as determined by the disability administration (see text for the description of the procedure for the determination of degree of disability). See the note to Table 4 for the demographic controls included in the regression.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

DI recipients. We find that our point estimates of benefit substitution in the entire sample are virtually identical whether or not we take the partners' responses into account. For men the benefit-substitution ratio becomes somewhat larger, while for women it becomes smaller, but neither difference is statistically significant. Though the increase in earnings crowd-out of 18 percentage points is not insubstantial in economic terms, it is statistically insignificant. Earnings responses of partners were previously studied by Cullen and Gruber (2000), who estimate that increased UI benefits paid to unemployed males are largely offset by their wives' decreased labor market earnings, as well as by Duggan, Rosenheck, and Singleton (2010), who exploit a policy change to show that an increase in disability benefits for Vietnam-era veterans caused a reduction in their spouses' labor supply. While partner responses could potentially be significant, and therefore are important to consider, we find that they play only a limited role in our setting. However, including the partner responses does decrease the precision of our estimates, which is why we exclude them from our other analyses.

The Dutch tax system makes no distinction between labor income and income from social assistance programs, and the effective marginal tax rate is roughly constant at a rate of approximately 40 percent. This implies that a €1,000 reduction in

TABLE 7—EARNINGS CROWD-OUT AND BENEFIT SHIFTING INCLUDING PARTNER RESPONSES

	Effect of reform per 1,000 €/year decrease in amount of original DI			<i>p</i> -value gender difference
	Full sample	Males	Females	
<i>Panel A. New DI benefit in 1999</i>				
Income from a new DI spell	0.008 (0.008)	0.001 (0.008)	0.046 (0.025)*	0.074
<i>Panel B. Other social assistance in 1999</i>				
Income from other social assistance	0.294 (0.082)***	0.308 (0.065)***	0.238 (0.317)	0.830
<i>Panel C. Labor market outcomes in 1999</i>				
Earnings	0.795 (0.209)***	0.718 (0.182)***	1.102 (0.749)	0.619
<i>Panel D. Total</i>				
Income except from original DI spell	1.097 (0.204)***	1.025 (0.185)***	1.386 (0.705)**	0.620

*Notes:* Standard errors are in parentheses. Each entry in the table comes from a separate IV regression based on the fuzzy RD design. The dependent variable is listed in the rows. Income and earnings are measured in thousands of euros per year. The variable that is instrumented (endogenous explanatory variable) is the amount of DI, so all coefficients can be interpreted as effect size per €1,000/year decrease in DI. The instrument itself is the treatment dummy (age less than 45 as of August 1, 1993). The regressions are based on 84,185 observations for the full sample, and on 55,772 and 28,413 observations for males and females, respectively. See the note to Table 4 for the demographic controls included in the regression.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

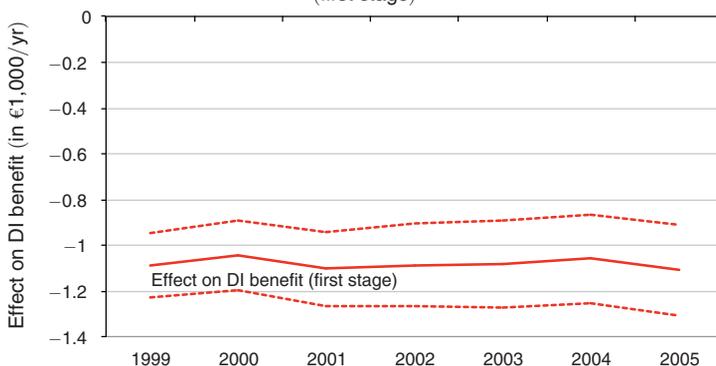
\* Significant at the 10 percent level.

DI benefits reduces government expenditures by about €600, net of taxes. According to the estimates in the first column of Table 7, a €1,000 reduction in DI benefits leads to increased expenditure on other social assistance programs (including reentry into DI) by €302 (= €294 + €8), which comes at a cost of  $0.60 \times 302 = €181$ , net of taxes. Table 7 also shows that a €1,000 reduction in DI benefits increases earnings by €795, which generates additional tax revenue of  $0.40 \times 795 = €318$ . Thus, the additional tax revenue from the labor supply response more than offsets the additional costs from social support substitution. Consequently, for each €1,000 (net €600) reduction in DI benefits, government spending decreases by  $600 + 318 - 181 = €737$ . Thus, the combined effect of the labor supply response and social support substitution increases government savings from the reduction in DI benefits by 23 percent ( $= (737 - 600)/600$ ).

#### F. Responses over Time

Responses to reductions in DI benefits could vary over time for a variety of reasons, including time limits on certain forms of social assistance and the search time required to find a suitable match in the labor market. Hence, focusing only on 1999 (about two years after the reexaminations for the relevant cohorts) yields an incomplete picture of the consequences of the reform. Another concern is that much of the decline in DI benefits for the under-45 cohort might be undone over

Panel A. Effect of more stringent reexamination on DI benefits in €1,000/year (first stage)



Panel B. Earnings crowd-out and benefit shifting using the 1999 first stage

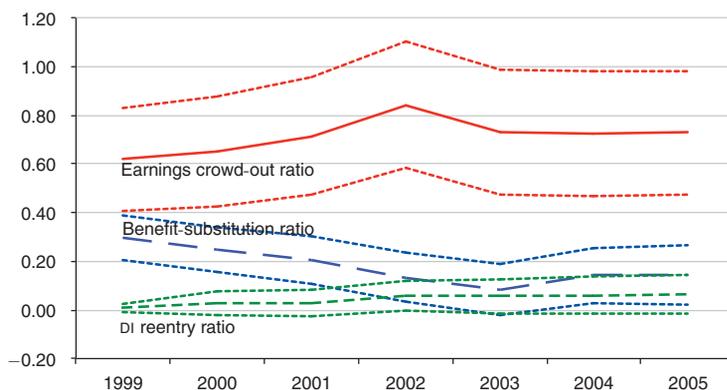


FIGURE 6. EARNINGS CROWD OUT AND BENEFIT SHIFTING OVER TIME

Note: The dotted lines represent the 95 percent confidence intervals.

time through “promotions” in the disability rating. We therefore repeated our main analyses for all years in our dataset.<sup>27</sup>

Panel A of Figure 6 presents estimates of the reduced-form RD regression of DI benefit amounts for each of the years from 1999 to 2005. This panel shows that the effect of the more stringent reexamination on DI benefit amounts is remarkably constant over time.<sup>28</sup> Panel B shows estimates of the benefit-substitution ratio and the earnings crowd-out ratio over time. We find that both the earnings crowd-out ratio and benefit-substitution ratio are positive and statistically significant in each year

<sup>27</sup> Additional DI reforms took place in 2002 and 2004. The first reform only affected new entrants, while the second reform led to a reexamination of people on DI who were younger than 50 on July 1, 2004. All individuals in our sample were older than 50 at that time. Thus, these reforms do not affect the individuals in our sample. However, the major overhaul of the DI system in 2006 did affect the individuals in our sample, which is why we end our sample period in 2005.

<sup>28</sup> We find a similar persistence of the effect of the more stringent reexamination when we examine the impact on disability ratings rather than the impact on benefit amounts. In online Appendix Table A9, we repeat the analysis of Table 3, but now for the years 1996 to 2005 (rather than 1996 to 1999, as in Table 3).

except 2003, when the benefit-substitution ratio is only marginally statistically significant. The degree to which individuals replace lost DI benefits with other forms of social assistance decreases over time, from 30 percent in 1999 to 14 percent in 2005. This decrease is statistically significant and is driven by the decreased reliance on UI benefits over time. Whereas UI benefits account for 20 percentage points of the benefit-substitution ratio in 1999, they only account for 5 percentage points in 2005. This decline is consistent with the fact that unemployment assistance is only available for a limited duration. This decline is partly offset by increased reliance on new DI spells. Income from new DI spells makes up for about 1 percent of lost DI benefits in 1999, but 6 percent of lost DI benefits in 2005. Reliance on General Assistance and other forms of social assistance remains roughly constant over time. The figure shows a slight increase over time in the earnings crowd-out ratio, which rises from 62 percent in 1999 to 71 percent in 2005, but this increase is not statistically significant. The fraction of earnings that comes from self-employment remains roughly constant over time.

#### IV. Conclusion

In this paper, we investigate the consequences of a reduction in the generosity of one social support program when that program is part of a larger system of social assistance programs. Especially in the case of social assistance for people in their prime age, there is relatively little evidence on the extent to which reduced generosity of one program induces increased labor supply, and to what extent it leads to more reliance on other social assistance programs instead. Examining the labor supply response of existing beneficiaries (as opposed to labor supply responses to qualify for a program) is important for policy because a response by the large stock of existing beneficiaries can quickly affect DI participation rates, whereas a response by new enrollees will only slowly affect overall DI participation. Showing a labor supply response of long-term DI beneficiaries (including individuals who are classified as fully disabled) also establishes that long-term participation in DI does not severely degrade one's labor market skills. Benefit substitution is of obvious policy relevance in many countries. While most existing studies have investigated spillover effects among programs for children or for people close to retirement, this paper examines benefit-substitution and earnings crowd-out effects for people on DI in their late forties. Finally, our paper recognizes that spillovers can be partly driven by responses from partners of people affected by the reform and that the spillovers may vary with the amount of time passed since the reform.

The combination of access to extensive administrative panel data and the presence of a cohort discontinuity in a reform law allows us to produce causal estimates of the effect of reduced DI generosity on participation in other social assistance programs. We find economically meaningful and statistically significant evidence of social support substitution. About two years after the implementation of the DI reform for our sample members, income drawn from other social assistance programs increases by €0.30 for each euro of reduced DI benefits. At 49 percent, the benefit substitution effect is especially pronounced for the fully disabled, whereas it is just 12 percent for partially disabled DI recipients. The benefit-substitution ratio

decreases over time and reaches 14 percent about eight years after implementation of the reform for our sample.

We also find a remarkable earnings rebound, given that all members of our sample were at least partially disabled and had, on average, been receiving DI for over a decade when the reform was implemented for our cohorts. On average, individuals were able to make up 62 percent of their foregone DI benefits through increased earnings, and this figure is similar for partially and fully disabled individuals. Between increased income from labor and other social assistance programs, individuals almost fully offset the decrease in DI benefits. Of course, these estimates are based on a relatively minor (10 percent on average) cut in DI benefits, and may not apply for larger cuts. Furthermore, because these estimates reveal average responses, they can also mask more severe impacts on total income for certain subgroups of DI recipients.

Benefit-substitution and earnings crowd-out estimates would obviously be different in different settings, but the direction in which the estimates would change is not clear. Our benefit-substitution figure may be higher than it would be in other countries, as the Netherlands has a relatively generous system of alternative social assistance programs. On the other hand, the reform we analyzed concerned a relatively minor reduction in DI generosity. Thus, many of those affected by the reform may not have qualified for means-tested alternative forms of social assistance. Moreover, alternative forms of social assistance may still have been less attractive than DI (despite the reduction in DI generosity).

While our specific coefficient estimates only directly apply to this particular Dutch DI reform, we believe our paper offers three general lessons that are widely applicable. First, our paper provides strong evidence that spillover effects between social assistance programs can be substantial, also for prime-aged individuals. Thus, any analysis of reforms of social assistance programs would be well advised to consider the possibility of benefit substitution. Second, we show that, among long-term disability recipients, there may still be a substantial capacity to change labor income in response to relatively moderate changes in DI generosity. In other words, labor supply among DI recipients is not just determined by limitations from disability, but also by economic incentives. Finally, our work emphasizes that it could be important in some cases to take into account the responses of the partners of the individuals directly affected by the reform and to consider the amount of benefit substitution and earnings crowd-out over the longer term.

## REFERENCES

- 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. 2008. "The Effect of Transfer Income on Labor Force Participation and Enrollment in Federal Benefits Programs: Evidence from the Veterans Disability Compensation Program." [http://www.law.yale.edu/documents/pdf/Intellectual\\_Life/D.Autor.pdf](http://www.law.yale.edu/documents/pdf/Intellectual_Life/D.Autor.pdf).
- Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell. 2011. "Heaping-Induced Bias in Regression-Discontinuity Designs." National Bureau of Economic (NBER) Working Paper 17408.
- 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: Dataset." *American Economic Journal: Economic Policy*. <http://dx.doi.org/10.1257/pol.6.4.34>.

- 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*, Vol. 3C, edited by Orley C. Ashenfelter and David Card, 3417–3528. Amsterdam: North Holland.
- Bovenberg, A. Lans.** 2000. "Reforming Social Insurance in the Netherlands." *International Tax and Public Finance* 7 (3): 345–68.
- Campolieti, Michele, and Chris Riddell.** 2012. "Disability Policy and the Labor Market: Evidence from a Natural Experiment in Canada, 1998–2006." *Journal of Public Economics* 96 (3–4): 306–16.
- Chen, Susan, and Wilbert van der Klaauw.** 2008. "The Work Disincentive Effects of the Disability Insurance Program in the 1990s." *Journal of Econometrics* 142 (2): 757–84.
- Coe, Norma B., and Kelly Haverstick.** 2010. "Measuring the Spillover to Disability Insurance Due to the Rise in the Full Retirement Age." Boston College Center for Retirement Research Working Paper 2010–21.
- Cullen, Julie Berry, and Jonathan Gruber.** 2000. "Does Unemployment Insurance Crowd Out Spousal Labor Supply?" *Journal of Labor Economics* 18 (3): 546–72.
- De Jong, Philip R., Maarten Lindeboom, and Bas van der Klaauw.** 2011. "Screening Disability Insurance Applications." *Journal of the European Economic Association* 9 (1): 106–29.
- Duggan, Mark G., and Melissa Schettini Kearney.** 2007. "The Impact of Child SSI Enrollment on Household Outcomes." *Journal of Policy Analysis and Management* 26 (4): 861–86.
- Duggan, Mark, Robert Rosenheck, and Perry Singleton.** 2010. "Federal Policy and the Rise in Disability Enrollment: Evidence from the Veterans Affairs' Disability Compensation Program." *Journal of Law and Economics* 53 (2): 379–98.
- Duggan, Mark G., Perry Singleton, and Jae Song.** 2007. "Aching to Retire? The Rise in the Full Retirement Age and Its Impact on the Disability Rolls." *Journal of Public Economics* 91 (7–8): 1327–50.
- French, Eric, and Jae Song.** 2011. "The Effect of Disability Insurance Receipt on Labor Supply." Federal Reserve Bank of Chicago Working Paper 2009–05.
- García-Gómez, Pilar, Hans-Martin von Gaudecker, and Maarten Lindeboom.** 2011. "Health, Disability and Work: Patterns for the Working Age Population." *International Tax and Public Finance* 18 (2): 146–65.
- Garrett, Bowen, and Sherry Glied.** 2000. "Does State AFDC Generosity Affect Child SSI Participation?" *Journal of Policy Analysis and Management* 19 (2): 275–95.
- Gruber, Jonathan.** 2000. "Disability Insurance Benefits and Labor Supply." *Journal of Political Economy* 108 (6): 1162–83.
- Gruber, Jonathan, and Jeffrey D. Kubik.** 1997. "Disability Insurance Rejection Rates and the Labor Supply of Older Workers." *Journal of Public Economics* 64 (1): 1–23.
- Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw.** 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (1): 201–09.
- Imbens, Guido, and Karthik Kalyanaraman.** 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* 79 (3): 933–59.
- Imbens, Guido W., and Thomas Lemieux.** 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.
- Inderbitzin, Lukas, Stefan Staubli, and Josef Zweimüller.** 2013. "Extended Unemployment Benefits and Early Retirement: Program Complementarity and Program Substitution." Institute for the Study of Labor (IZA) Discussion Paper 7330.
- 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): 2071–82.
- Koning, Pierre W. C., and Daniel J. van Vuuren.** 2010. "Disability Insurance and Unemployment Insurance as Substitute Pathways." *Applied Economics* 42 (5): 575–88.
- Kostøl, Andreas R., and Magne Mogstad.** 2014. "How Financial Incentives Induce Disability Insurance Recipients to Return to Work." *American Economic Review* 104 (2): 624–55.
- Kubik, Jeffrey D.** 2003. "Fiscal Federalism and Welfare Policy: The Role of States in the Growth of Child SSI." *National Tax Journal* 56 (1): 61–79.
- Lammers, Marloes, Hans Bloemen, and Stefan Hochguertel.** 2013. "Job Search Requirements for Older Unemployed: Transitions to Employment, Early Retirement and Disability Benefits." *European Economic Review* 58 (2): 31–57.
- Lee, David S., and Thomas Lemieux.** 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355.

- Li, Xiaoyan, and Nicole Maestas.** 2008. "Does the Rise in the Full Retirement Age Encourage Disability Benefits Applications? Evidence from the Health and Retirement Study." Michigan Retirement Research Center (MRRRC) Working Paper 2008–198.
- Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand.** 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review* 103 (5): 1797–1829.
- Maestas, Nicole, and Jae Song.** 2011. "The Labor Supply Effects of Disability Insurance: Evidence from Automatic Conversion Using Administrative Data." Michigan Retirement Research Center (MRRRC) Working Paper 2010–247.
- Marie, Olivier, and Judit Vall Castello.** 2012. "Measuring the (Income) Effect of Disability Insurance Generosity on Labour Market Participation." *Journal of Public Economics* 96 (1–2): 198–210.
- 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.** 2012. "Disability Insurance Receipt and Changes in Health and Human Capital." Unpublished.
- Parsons, Donald O.** 1980. "The Decline in Male Labor Force Participation." *Journal of Political Economy* 88 (1): 117–34.
- Schmidt, Lucie, and Purvi Sevak.** 2004. "AFDC, SSI, and Welfare Reform Aggressiveness: Caseload Reductions vs. Caseload Shifting." *Journal of Human Resources* 39 (3): 792–812.
- Singleton, Perry.** 2012. "Earnings of Rejected Applicants to the Social Security Disability Insurance Program." *Economics Letters* 116 (2): 147–50.
- Staubli, Stefan.** 2011. "The Impact of Stricter Criteria for Disability Insurance on Labor Force Participation." *Journal of Public Economics* 95 (9–10): 1223–36.
- Uitvoeringsinstituut Werknemersverzekeringen (UWV).** 2006. *Kroniek van de Sociale Verzekeringen 2006–Wetgeving en Volume-Ontwikkeling in Historisch Perspectief*. UWV: Amsterdam.
- von Wachter, Till, Jae Song, and Joyce Manchester.** 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.