

How Financial Incentives Induce Disability Insurance Recipients to Return to Work*

Andreas Ravndal Kostøl[†] Magne Mogstad[‡]

Forthcoming in the American Economic Review

Abstract

Using a local randomized experiment that arises from a sharp discontinuity in Disability Insurance (DI) policy in Norway, we provide transparent and credible identification of how financial incentives induce DI recipients to return to work. We find that many DI recipients have considerable capacity to work that can be effectively induced by providing financial work incentives. We further show that providing work incentives to DI recipients may both increase their disposable income and reduce program costs. Our findings also suggest that targeted policies may be the most effective in encouraging DI recipients to return to work.

*We thank three anonymous referees, Rolf Aaberge, Richard Blundell, Espen Bratberg, Christian Brinch, Paul Devereux, Per Johansson, Tarjei Havnes, Torfinn Harding, Katrine Løken, Knut Røed, Kjetil Telle, Kjell Vaage, Till Von Wachter, and a number of seminar and conference participants for useful comments and suggestions. This research was supported by the U.S. Social Security Administration (SSA) through grant #1 DRC12000002-01-00 to the National Bureau of Economic Research (NBER) as part of the SSA Disability Research Consortium. The findings and conclusions expressed are solely those of the author(s) and do not represent the views of SSA, any agency of the Federal Government, or the NBER.

[†]Research Department, Statistics Norway, P.O. Box 8131 Dep. NO-0033 Oslo, Norway; University of Bergen. E-mail: andreas.kostol@ssb.no

[‡]Department of Economics, University College London, Gower Street, London WC1E 6BT, UK; Statistics Norway. E-mail: m.mogstad@ucl.ac.uk

Keywords: Disability insurance, financial incentives, labor supply, regression discontinuity design

JEL-codes: H53, H55, I18, J21

Disability Insurance (DI) programs have long been criticized by economists for apparent work disincentives. Some countries have recently modified their program such that DI recipients are allowed to keep some of their benefits if they return to work, and other countries are considering similar return-to-work policies. For example, the United Kingdom recently introduced the Pathways-to-Work program. One component of the reform was an increase in the financial incentives meant to encourage DI recipients to return to work; they were allowed to keep approximately 50% of their disability benefits for up to 12 months. Another example is the proposed change in the U.S. DI program, known as the “\$1 for \$2 offset”. Under this policy, a DI recipient’s benefits would be reduced by \$1 for every \$2 of the person’s earnings above the substantial gainful activity (SGA) threshold.

Advocates of such return-to-work policies claim that they increase the employment rate among DI recipients and even encourage some to exit the DI rolls entirely; thus, they argue, the welfare of DI recipients can be enhanced, and the cost of DI programs can be reduced. A counterargument is that allowing DI recipients the option to return to work and keep a fraction of their benefits makes the program more generous, both for potential applicants and for current recipients; in theory, therefore, return-to-work policies could reduce labor supply, lead to fewer exits from DI by current recipients, and encourage more applications for DI.¹

The challenge in assessing these arguments is that the evidence base for return-to-work policies for DI recipients is scarce. In 1999, President Clinton signed a federal law that required the Social Security Administration to undertake a randomized controlled trial. The study was supposed to determine the magnitude of the labor supply response that would likely occur if a policy change was made that allowed DI recipients to keep some portion of their benefits if they returned to work. However, no such experiment was conducted, and there is little non-experimental evidence

¹These arguments are presented in several of the bills that have been brought to the floor of the U.S. Congress, outlining options for reforming the work incentives of the SSDI program, including the Work Incentive and Self Sufficiency Act of 1996, the Rehabilitation and Return to Work Opportunity Act of 1996, and the Transition to Work Act of 1997 (see e.g. Stapleton et al., 2008; Livermore, Roche and Prenovitz, 2009; Autor, 2011).

to guide policy makers.

This paper helps to close that gap by analyzing the consequences of providing financial incentives to DI recipients to encourage them to return to work. In January 2005, the Norwegian government introduced such a program: the benefits of DI recipients would be reduced by approximately \$0.6 for every \$1 in earnings that they accumulated above the SGA threshold. However, only recipients who had been awarded DI before January 1 of 2004 were eligible for the return-to-work program. Because the cut-off date for eligibility was set retroactively, individuals were unable to gain entry to the program by manipulating the award date. This sharp discontinuity in DI policy therefore provides an attractive setting to analyze the impact of financial incentives using a regression discontinuity (RD) design that compares recipients awarded DI just before January 1 of 2004 (i.e., the treatment group) and just after that date (i.e., the control group).

The insights from our RD analysis may be summarized with three broad conclusions. First, many DI recipients have a considerable capacity to work that can be effectively induced by providing financial work incentives. Three years after implementation, the return-to-work program has increased the labor force participation of DI recipients aged 18-49 by 8.5 percentage points. In comparison, 3.4 percent of the control group participated in the labor force, and the labor force participation among rejected DI applicants of the same age was just above 30 percent. Second, the return-to-work program produced a substantial increase in the earnings of DI recipients. These gains in earnings were accompanied by a small increase in disposable income and a substantial reduction in the cost of the program. The reduction in costs is attributable to a significant decrease in benefits and a small increase in the taxes paid by DI recipients. Third, there is significant heterogeneity in the responsiveness to financial work incentives, indicating that targeted policies may be most effective in inducing DI recipients to return to work. Among DI recipients aged 50-61, who are approaching the retirement age, there is no evidence of any impact of the program. Within the group of DI recipients aged 18-49, the estimates vary substantially, with the strongest responses to financial incentives among males, individuals with high education levels or more labor market

experience, and people in areas with low unemployment.

To assess the internal validity of our RD design, we perform a number of specification checks, all of which support our main results. We also take several steps to shed light on the generalizability of our findings. We begin by describing similarities and differences between the DI systems in the U.S. and in Norway. Next, we use the approach proposed by Bound (1989) to assess the work capacity and labor supply elasticity of DI recipients. Bound uses the labor force participation rate for rejected DI applicants as an upper bound estimate of the labor force participation rate for DI recipients if the DI program had not existed. When applying this approach to the Norwegian data, we obtain upper bound estimates that are quite similar to the recent U.S. evidence. This indicates that the work capacity and labor supply elasticity of DI recipients in the U.S. are comparable to those in Norway, lending support to the external validity of our analysis of the return-to-work program.

Because the cut-off date for eligibility to the return-to-work program was set retroactively, we are not able to estimate the level of induced entry that may occur when DI recipients are given the option to return to work and keep a fraction of their benefits. We do, however, make a calculation for the size of the induced entry that would have to be generated by the return-to-work policy to lead to an increase in program costs. Our calculation indicates that the elasticity of entry response to the change in DI benefits needs to be much higher than what is suggested by the existing literature (see e.g. Gruber, 2000; Campolieti, 2004).

Our paper is primarily related to a fairly small literature on the consequences of providing financial incentives that encourage DI recipients to return to work. Hoynes and Moffitt (1999) conclude via numerical simulations that some of the proposed reforms intended to incentivize work among DI recipients are unlikely to be successful. Benitez-Silva, Buchinsky and Rust (2010) use a calibrated life-cycle model to forecast the effects of the “\$1 for \$2 offset” policy. These authors find that depending on the stigma costs associated with DI, the policy may encourage work by DI recipients but could also induce entry into the DI program. Another related strand of the litera-

ture examines return-to-work reforms that do not (exclusively) involve changes in financial work incentives. The results are mixed. The medical reassessment of beneficiaries through continuity disability reviews and trial work periods via the Ticket to Work program does not appear to have encouraged many people to return to work in the U.S. (Autor and Duggan, 2006). However, several other policies appear to be quite effective in inducing return to work on the part of disability beneficiaries.²

Our paper is also related to a growing literature on the causes and consequences of the growth in DI rolls (for a review, see Autor and Duggan (2006)). Such growth poses significant risks to the finances of DI programs and Social Security systems more generally. It also raises troubling questions regarding whether the programs are being misused by recipients. We show that many individuals who are deemed totally and permanently disabled by the Social Security Administration indeed have considerable work capacity and that their labor supply is fairly elastic with respect to changes in financial work incentives.

Our findings complement the current evidence on the work capacity and labor supply elasticity of DI recipients. Some of the evidence comes from studies that examine the variation in benefit generosity across regions and over time to estimate the relationship between disability benefits and labor force participation. For example, Gruber (2000) studies the policy change conducted in Canada. He estimates the elasticity of labor force non-participation with respect to DI benefit levels in the range of 0.28 and 0.36. Another example is Autor and Duggan (2003), who use differential time variation in average benefits across regions to identify the impact of DI on the labor force participation of low-skilled workers. The authors conclude that the DI system has provided many low-skilled workers with a viable alternative to unemployment.³

Much of the remaining evidence comes from studies that use the labor supply of rejected dis-

²See e.g. Adam, Bozio and Emmerson (2011), Moore (2011), Borghans, Gielen and Luttmer (2012), and Campolieti and Riddell (2012).

³There is also a considerable body of evidence of how local labor market conditions and wages affect entry to DI (see e.g. Bratberg, 1999, Black, Daniel and Sanders, 2002, Rege, Telle and Votruba, 2009, Bratsberg, Fevang and Roed, 2010).

ability applicants to estimate the labor supply of DI recipients if the DI program had not existed. Bound (1989) finds that the labor force participation rate of DI recipients in the U.S. would have been at most 30 percentage points higher had they not received disability benefits. Recent studies have extended this analysis, in part by analyzing different subgroups and different time periods but also with quasi-experimental variation in the disability determination process to improve the comparability of DI recipients and rejected DI applicants.⁴

The remainder of the paper proceeds as follows. Section I reviews the key facts regarding the DI program in Norway, compares the system with that of the U.S., and discusses the return-to-work program and its expected impact. Section II describes the data and discusses the empirical strategy, Section III presents the empirical results, and Section IV concludes.

I Background

A The Norwegian DI program

In Norway, DI benefits are designed to provide partial earnings replacement to all workers under the full retirement age who are unable to engage in substantial gainful activity because of a medically determinable physical or mental impairment that has lasted for at least a year. The program is part of the broader Social Security System and is financed by employer- and employee-paid taxes.⁵

Disability determination process

To apply for DI benefits, an individual must submit an application to a Social Security Administration (SSA) office. The disability examiners first review the non-medical criteria. Only adults

⁴See Chen and van der Klaauw (2008), French and Song (2009), von Wachter, Song and Manchester (2011), and Maestas, Mullen and Strand (2012).

⁵Throughout this paper, we focus on the system for individuals who are deemed totally and permanently disabled by the Social Security Administration. This category includes the majority of DI recipients. We have chosen this focus because the return-to-work program was not intended to target the partially or temporarily disabled.

below the (mandatory) retirement age of 67 years are eligible for DI. If the applicant meets the non-medical criteria, disability examiners and medical staff assess medical evidence from one or more of the applicant's health care providers regarding the applicant's ability to perform work-related activities. If the applicant has an impairment that is undoubtedly disabling, a disability award is made. In other cases, the next step is to consider the applicant's overall ability to work, taking into account his or her health status, age, education, and work experience as well as the transferability of the applicant's skills. If the disability examiner concludes that the applicant cannot be expected to engage in any substantial gainful activity, a disability award is made.

The level and duration of DI benefits

The level of DI benefits received is determined using a formula for which the first step is to calculate the average indexed annual earnings (AIE). Past earnings are indexed to the present value using an inflator that is equal to the average wage growth in the economy. The years with the lowest earnings are excluded from these calculations. The proportion of income that is replaced decreases as past earnings increase so that low-wage workers replace a larger fraction of their earnings as compared to high-wage workers.⁶

Once a person is enrolled in the DI program, there are three main pathways out of the program. First, the DI recipient may reach the retirement age, at which point he or she will shift to the old-age pension program of Social Security. In 2003, 74 percent of those who left DI did so for this reason. Second, the recipient may die. Death accounted for 19 percent of all program exits in 2003. Third, the recipient may no longer meet the non-medical or medical criteria for receiving DI benefits. For example, an individual may return to work and earn more than the permitted maximum amount. In 2003, eligibility-based exits accounted for only 7 percent of all exits from the program.

⁶Consider, for example, the population of individuals awarded DI in 2003. If we divide this population into three equally sized groups according to their earnings in 2000-2002, the average replacement rate is 90 percent in the bottom tertile, 55 percent in the middle tertile, and 50 percent in the top tertile. See Rege, Telle and Votruba (2009) for a detailed description of the formula determining the DI benefits.

B The return-to-work program and its expected effects

In January 2005, the Norwegian government introduced a program to encourage DI recipients to return to work. However, only recipients who had been awarded DI before January 1 of 2004 were eligible for the return-to-work program (i.e. the treatment group). Recipients who had been awarded DI after January 1 of 2004 had to abide by the current rules (i.e. the control group). Below, we describe the work incentives in the two programs and use static labor supply theory to make economic predictions.⁷

The programs

There is only one difference between the current rules and the return-to-work program: the manner in which benefits are reduced if earnings exceed the SGA threshold. In both programs, there is a 12-month waiting period after DI is awarded. If the DI recipient is working during the waiting period, he or she becomes ineligible for DI benefits. After the waiting period, both programs have a trial working period allowing the DI recipient to test his or her work capacity without impacting future benefit payments. Specifically, a person is allowed to have earnings above the SGA threshold for up to five consecutive years without risking reduced benefits should that person's earnings fall below SGA after these five years. If a person has earnings above SGA for more than five consecutive years, benefits would no longer be provided in full should the person's earnings fall below SGA in the future.

Both programs specify benefits as functions of earnings, and not hours. During the trial working period, the formulas for annual DI benefits in the return-to-work program (B^T) and the current rules (B^C) depend on annual earnings (W) and the maximum annual benefit level (B^0) in the following way:

⁷The description of the two programs is based on the Norwegian government's white paper from 2007 on disability insurance (NOU, 2007). Throughout this paper, all monetary figures are fixed at 2010 level. For the figures expressed in U.S. dollars (USD), we have used the following exchange rate: NOK/USD = 6.

$$B^C = \begin{cases} B^0 & \text{if } W \leq SGA \\ & \text{where } SGA = \$12,450 \text{ p.a.} \\ B^0 - \tau(W - SGA) - \theta & \text{if } SGA < W \leq E \\ & \text{where } \theta = \frac{SGA}{AIE} B^0, \tau = \frac{B^0}{AIE} \\ 0 & \text{if } W > E \\ & \text{where } E = 0.8 \times AIE \end{cases} \quad B^T = \begin{cases} B^0 & \text{if } W \leq SGA \\ & \text{where } SGA = \$12,450 \text{ p.a.} \\ B^0 - \tau(W - SGA) & \text{if } SGA < W \leq e \\ & \text{where } \tau = \frac{B^0}{AIE} \\ 0 & \text{if } W > e \\ & \text{where } e = 0.8 \times (SGA + AIE) \end{cases}$$

Benefits are provided in full if the person's annual earnings are less than SGA threshold, which is set at monthly earnings of about USD 1,050. If earnings exceed SGA, the two programs differ. Current rules create a discontinuous fall θ in benefits at the first dollar earned above SGA; after this notch, there is a τ reduction in benefits for every dollar in earnings; benefits are provided until earnings exceed the maximum permitted amount E ; this situation implies the absence of disability and therefore signals the individual's ineligibility for DI benefits. The return-to-work program eliminates the discontinuous fall in benefits at the SGA threshold; it also increases the maximum permitted amount to e .

In our analysis, the waiting period implies that members of both the treatment group and the control group would become ineligible for DI if they were working in 2004. The return-to-work program takes effect on January 1, 2005. From that date onwards, members of the treatment group are incentivized to work by allowing them to keep more of their benefits if earnings exceed SGA, as compared to members of the control group. Our empirical analysis compares the behavior of the treatment group and the control group over the period 2005-2007. Due to the trial working period, members in both groups could have annual earnings above SGA in these years without risking reduced benefits should their earnings fall below the SGA in the future (e.g. 2008 or 2009).

The budget constraints

The top graph in Figure 1 illustrates the difference in work incentives across the treatment and the control group. This figure translates the benefit formulas into stylized budget constraints in the income-leisure space. We consider the wage and benefit levels associated with a typical DI recipient. To compute the benefit levels, we use the average work history of recipients awarded DI during the period between July 2003 and February 2004. We set the hourly wage equal to the average hourly wage in a sample of workers who are not on DI but have observable characteristics similar to those of the DI recipients.⁸ The line AH (slope equal to the hourly wage rate) applies in the absence of any DI program. For the sake of simplicity and with minimal loss of generality about the expected impacts of the return-to-work program, we disregard income taxation and dependent benefits.

Insert Figure 1 approximately here

The budget constraint under the current rules is represented by $BCDEFH$. Benefits are provided in full at budget segment BC . At 10 hours of work per week, the person's earnings are above the SGA threshold, and the marginal tax rate on another hour of work exceeds 600 percent, thus creating a notch in the budget constraint (represented by the line CD). At the budget segment DE , hours of work range from 10 to 32 per week, and the person's benefits are phased out at a rate of approximately \$0.6 for every \$1 in earnings. At 32 hours of work per week, earnings exceed the maximum permitted amount and the person becomes ineligible for DI benefits. This creates a second notch that is given by the line EF .

The budget constraint in the return-to-work program is $BCefH$. The program eliminates the first notch in the budget constraint under current rules and phases out benefits at budget segment Ce

⁸To compute the hourly wages, we use information from the Wage Statistics Survey in 2005. On the basis of this sample of workers (who are not on DI), we regress hourly wages on the covariates included in Table 1 (except diagnoses) and a full set of indicators for municipality of residence. We allow for a flexible functional form by including a large set of dummies for different covariate values. Using the estimated wage equation, we predict hourly wages for individuals on DI, and compute the average wage rate.

at a rate of approximately \$0.6 for every \$1 in earnings. The program also increases the maximum permitted amount so that the individual can work as much as 43 hours per week before he or she becomes ineligible for DI benefits. This shifts the second notch to the left; it is now given by the line ef .

The expected impacts

What are the expected effects of the transformation of the budget constraint from the current rules to the return-to-work program? To answer this question, we consider a static labor supply model where DI recipients choose hours of work at the given offered wage and offered wages are constant; we assume that preferences are convex and that leisure and income are normal goods. The idea is to compare the labor supply incentives for someone facing the current rules to the counterfactual state in which the person is assigned to the return-to-work program. As in Figure 1, we consider the wage and benefit levels associated with a typical DI recipient.⁹

We first focus on the case in which DI recipients can freely choose hours of work. Consider a DI recipient who under current rules would locate at point C , working 10 hours and receiving the maximum benefits. Depending on the individual's preferences, assignment to the return-to-work program could lead to either of two outcomes. First, the individual might continue to work 10 hours and receive the maximum benefit with no change in income. Second, he or she might enter the labor market, moving from C to some point at the budget segment Ce ; benefits are gradually phased out, while income rises. Consider next an individual who under current rules would locate at points on the budget segment DE that give higher income than at point C . Then we would expect the increase in benefits accompanying assignment to the return-to-work program to reduce hours of work and increase total income.¹⁰ Now imagine that an individual's preferences are such that

⁹The work incentives depend on the wage rate the DI recipients face. In particular, if the wage rate is lower (higher) than the wage of the typical DI recipients, the notches in the budget constraints at C , E and e are associated with more (fewer) hours of work. Independently of the wage rate, however, the static labor supply model predicts that the return-to-work program will increase the population share with earnings above SGA.

¹⁰If the individual would locate at point E under current rules, then the positive substitution effect at this notch could

full-time while receiving DI benefits.

In sum, the expected impacts of the return-to-work program are heterogeneous and may be negative, zero or positive. The static labor supply model predicts that the return-to-work program will increase the population share with earnings above SGA, independently of whether DI recipients freely choose hours of work. There are, however, no unambiguous predictions about the estimated mean impacts of the return-to-work program: the sign and magnitude depend on the size of the different responses weighted by the relative number of DI recipients along the budget constraint under current rules. As discussed in more detail below, there are few DI recipients who earn more than SGA under the current rules. It is therefore likely that the return-to-work program will increase average earnings and income, and that it will reduce average benefits. We will nevertheless complement the usual mean impact analysis with estimates of the probability that earnings fall in different segments of the budget constraint.

C Comparison with the U.S. program

The DI programs in Norway and the U.S. are quite similar and feature only three important differences.¹¹

The first difference is that the incidence of receipt of DI benefits is lower in the U.S. than in Norway. Figure 2 shows this distinction by displaying the evolution of DI in the two countries. Whereas the rate of DI receipt in a given year is consistently higher in Norway than in the U.S.,¹² the time trends are quite similar. From 1961 to 2005, the rate of receipt increased from 2.2 to 10.4

¹¹Our discussion of the U.S. system draws primarily on Autor and Duggan (2006), and pertains only to the SSDI program. More than 80 percent of non-elderly U.S. adults are insured against the risk of disabling physical or mental illness by SSDI (Autor and Duggan, 2006).

¹²Unlike the Norwegian DI program, the U.S. SSDI program does not offer temporary disability benefits. To improve comparability, Figure 2 therefore excludes temporarily disabled individuals when computing the number of DI recipients in Norway. The cross-country difference in coverage by the DI program is unlikely to explain the entire discrepancy in the incidence of DI: although virtually all non-elderly adults are covered in Norway, more than 80 percent of all non-elderly adults are covered in the U.S. It is unclear whether the remaining difference is a function of underlying differences in screening stringency, the frequency with which people apply for disability benefits, or the health of the population. However, Milligan and Wise (2011) argue that differences in health are unlikely to explain much of the observed differences in rates of DI receipt across developed countries.

percent in Norway and from 0.8 to 4.2 percent in the U.S. In both countries, the expansion of the DI rolls in recent decades appears to be driven by the liberalization of the screening process, which led to a rapid increase in the share of DI recipients suffering from difficult-to-verify disorders such as mental illness and musculoskeletal disease.¹³ Because these are early-onset disorders with low age-specific mortality, DI recipients with such diagnoses tend to participate in the program for relatively long periods. As a result, the DI exit rates have decreased in the last few decades. In 1985, the DI exit rate was approximately 12.1 percent in the U.S. and 10.4 percent in Norway. In both countries, this rate has trended steadily downward since that time and reached approximately 7 percent in 2004. As shown in Figure 2, this decline has been driven both by a decrease in the fraction of DI recipients who reach the full retirement age and by a decrease in the fraction of DI recipients who die.

Insert Figure 2 approximately here

The second difference between the two programs is that the U.S. SSDI program is less generous. Appendix Figure A.1 shows this difference by presenting stylized budget constraints in the income-leisure space for the two DI programs. To compute the budget constraint for the U.S. program, we use the same work history and wages as in Figure 1. We see that the maximum benefit level received if one's earnings are below the SGA threshold is lower in the U.S. than in Norway; for the worker we consider, the replacement rate would be 31 percent in the U.S. program and 58 percent in the Norwegian program.¹⁴ Although the two programs have similar SGA thresholds (USD 1,050 per month in Norway and USD 1,000 per month in the U.S.), the U.S. program takes earnings above the SGA threshold to imply the absence of disability and therefore *ineligibility* for

¹³See Autor and Duggan (2006) for a discussion of this phenomenon. In the U.S., the 1984 congressional reforms shifted the focus of screening from medical to functional criteria. In Norway, the medical eligibility criteria were relaxed earlier and more gradually.

¹⁴For the sake of simplicity, our calculations for the replacement rates disregard income taxation, dependent benefits, and health insurance. In both countries, the DI programs provide dependent benefits. In addition, DI recipients in the U.S. receive health insurance coverage through the federal Medicare program, which increases the replacement rate somewhat. In Norway, all citizens are eligible for health insurance through the Social Insurance System.

DI; this increases the tax rate associated with the first notch and eliminates the second notch, as compared to the Norwegian program. Figure A.1 also compares the budget constraints associated with the proposed “\$1 for \$2 offset” policy with the return-to-work program. It is clear that the “\$1 for \$2 offset” policy allows DI recipients who return to work to retain slightly more of their benefits. We also see that the “\$1 for \$2 offset” policy does not specify a maximum permitted amount, which created the second notch in the budget constraints in Norway.

The last difference is that DI recipients in Norway tend to be somewhat older and to have slightly higher earnings prior to disability award. Appendix Table A.1 shows this distinction by reporting the key characteristics of DI recipients in the U.S. and in Norway. The differences in the population characteristics are, however, less pronounced than one might expect given the differences in the generosity of the program. For instance, the majority of DI recipients suffer from difficult-to-verify disorders including mental illness and musculoskeletal disease in both the U.S. (57.4 %) and Norway (61.4 %).

II Data and empirical strategy

A Data and sample selection

Our analysis employs several data sources that we can link through unique identifiers for each individual. The data on DI come from social security registers that contain complete records for all individuals who entered the DI program during the period 2001-2005. The data set contains information on the month when DI was awarded (or denied), the individual’s medical diagnosis and average indexed earnings, and the level of DI benefits received. We link this data with administrative registers provided by Statistics Norway, using a rich longitudinal database that covers every resident from 1967 to 2007. For each year, it contains individual demographic information (regarding sex, age, and number of children), socio-economic data (regarding years of education,

earnings, cash transfers, taxes, and private vs. public sector employment), and geographical identifiers that indicate place of residence.

The outcomes that we consider are defined as follows. The first key outcome is labor force participation. As in Maestas, Mullen and Strand (2012), labor force participation is an indicator variable that is equal to one if annual earnings exceed the annual SGA threshold, set at approximately USD 12,500 per year. Unfortunately, we are unable to measure labor supply at the intensive margin because we do not have data on working hours.¹⁵ The second key outcome is annual gross earnings, which includes wages and income from self-employment. We also consider the effect of the program on disposable income, incorporating earnings and benefits net of taxes. In addition, we examine the impact of the program on DI benefits received and taxes paid. The last outcome that we consider is exit from the DI program, which occurs if an individual is classified by the Social Security Administration as no longer eligible for DI.

Our empirical analysis considers individuals aged 18-61 at the time of the DI award. The reason for the upper age restriction is that Norwegian workers can receive early retirement benefits beginning at age 62; the lower age restriction is because individuals under 18 are not eligible for DI. Throughout the empirical analysis, we partition the sample into two groups according to age. One reason is that the cost of working may be different for younger DI recipients, who more often suffer from difficult-to-verify disorders such as mental illness (Autor and Duggan, 2006; von Wachter, Song and Manchester, 2011). Another reason is that younger DI recipients may gain more from returning to the labor force because they have a longer horizon until retirement (Bound, Burkhauser and Nichols, 2003; von Wachter, Song and Manchester, 2011). In our main analysis, we will focus on individuals who were aged 18-49 when they were awarded DI. In subsection III.F, we also report results for older DI recipients, those who were aged 50-61 when they were awarded DI.

¹⁵Figure 1 uses the Wage Statistics Survey to represent the distribution of working hours for a sample of workers who are not on DI. Unfortunately, the sample provided by the Wage Statistics Survey is too small for us to analyze the impact of the return-to-work program on hours of work.

To assess the impact of the return-to-work program, we make two notable sample restrictions. First, we restrict the sample to individuals who were deemed totally and permanently disabled by the SSA, because the return-to-work program was not targeted at partially or temporarily disabled individuals. Second, we exclude individuals who (prior to DI award) resided in a county with a wage-subsidy policy. Five (of nineteen) counties provided wage subsidies to firms that hired DI recipients. The reason for this sample restriction is to avoid confounding the effects of the return-to-work program with the impact of the wage policy. This gives a total of 9,471 individuals aged 18-61 who were awarded DI in 2003 (2,896 individuals were between the ages of 18 and 49, while 6,575 individuals were age 50-61). In the baseline RD specification, the estimation sample consists of individuals who were awarded DI between November 2003 and February 2004 (897 individuals aged 18-49 and 2131 individuals aged 50-61).

B RD design

In our sharp RD design, assignment to the return-to-work program is a deterministic function of the assignment variable, the date of the DI award (X): only recipients who had been awarded DI before January 1 of 2004 were eligible for the return-to-work program. The RD design uses separate regressions on each side of this cut-off date (c). The regression model for the treatment group is applied to the left side of the cut-off date ($X < c$)

$$(1) \quad Y = \alpha_l + f_l(c - X) + \varepsilon_l,$$

whereas the regression model for the control group is applied the right side of the cut-off date ($X > c$)

$$(2) \quad Y = \alpha_r + f_r(X - c) + \varepsilon_r,$$

where f_r and f_l are unknown functional forms. The RD estimate of the return-to-work program is then given by the difference between the estimated regression intercepts on the two sides of the cut-off date

$$(3) \quad \hat{\tau}^{RD} = \hat{\alpha}_l - \hat{\alpha}_r.$$

To make inferences about the impact of the program, we pool the two regressions and follow Fuji, Imbens and Kalyanaraman (2009) in calculating the standard errors robust to heteroscedasticity.

To implement the RD design, we need to specify f_r and f_l and decide on the window on each side of the cut-off date. Our first specification uses a local linear regression with triangular kernel density and 2 months of bandwidth on each side of the cut-off date. This choice of bandwidth corresponds to the optimal bandwidth we obtain when following Imbens and Kalyanaraman (2012). Our second specification uses a window of just one month on each side of the cut-off date. Because we have monthly data for DI awards, the RD model is then equivalent to a first-difference (FD) model: the average outcome for recipients who were awarded DI in December 2003 is compared with the average outcome for recipients who were awarded DI in January 2004.

C Graphical evidence

A virtue of the RD design is that it provides a transparent way of showing how the program impact is identified. To this end, we begin with a graphical depiction of how the labor market outcomes, the covariates, and the density of the assignment variable vary around the cut-off date, before turning to a more detailed regression-based analysis.

The top panel of Figure 3 shows the mean labor force participation and average earnings for recipients who were awarded DI between July 2003 and June 2004; the bottom panel of Figure 3 shows the means for the same labor market outcomes for recipients who were awarded DI between July 2002 and June 2003; and Figure 4 graphs the covariates for recipients who were awarded DI

between July 2003 and June 2004. In each figure, we plot the unrestricted monthly means and the estimated monthly means using local linear regression applied to each side of the cut-off dates. Whereas the regression lines better illustrate the trends in the data and the size of the jumps at the cut-off, the unrestricted means indicate the underlying noise in the data. In each graph, the scale of the y-axis is set equal to ± 0.5 standard deviation of the respective variable. By standardizing the y-axes in this way, it becomes easier to compare the trends in the data and the sizes of the jumps at the cut-off dates across the graphs.

Figure 3 shows evidence of a discontinuity in each outcome around January 1 of 2004, suggesting that the return-to-work program had substantial effects on the labor supply of DI recipients. In comparison, there is no sign of discontinuous changes in labor market outcomes around January 1 of 2003; this finding is reassuring because there were no changes in assignment to the return-to-work program around this cut-off date.

Insert Figure 3 and Figure 4 approximately here

Note that the gray shaded areas in Figures 3 and 4 mark the months in which DI recipients were affected by a potentially confounding policy change. To address this issue, we will restrict the control group to the non-shaded area throughout our empirical analysis: that is, the control group will only include individuals who were awarded DI in January and February 2004. The policy change of concern was the introduction of temporary DI benefits in January 2004. However, individuals who had submitted their DI applications before January 2004 were ineligible for temporary DI. Because it generally takes between 3 and 8 months to process the application, recipients who were awarded DI before March 2004 were not affected by the introduction of temporary DI benefits.¹⁶ And because the cut-off date for eligibility to temporary DI was set retroactively, individuals were unable to gain or avoid entry to temporary DI by manipulating the application date.¹⁷

¹⁶The guidelines of the SSA state that applicants should expect that it takes 3-8 months (and not more than 12 months) to process the initial application. For example, for individuals who applied for DI in December 2003, the reported average processing time was 176 days.

¹⁷Our results are qualitatively the same and quantitatively similar if we extend the control group to include individ-

As a consequence of the individuals being unable to manipulate the award date, any pre-determined characteristic should have the same distribution just before and just after January 1 of 2004. Figure 4 supports this conjecture, showing that the sample mean of the (pre-determined) covariates tend to be quite stable around the cut-off date. One exception is the DI rate of the municipality of residence (prior to DI benefit award), which fluctuates somewhat over time and decreases around the cut-off date. However, the decrease in the local DI rate is statistically insignificant at conventional levels; and based on the large number of covariates and pre-reform outcomes we consider, the probability of observing changes around the cut-off date by chance alone is quite large. It is, however, reassuring that there is little if any correlation between the local DI rate and the outcome variables.¹⁸ As a consequence, controlling for the local DI rate barely moves our estimates.

Another implication of local randomization is that the aggregate distribution of the assignment variable, the date of the DI award, should be continuous around the cut-off date; an increase in the density would indicate sorting around the threshold, calling into question the appropriateness of our RD design. In Appendix Figure A.2, we examine the density of the assignment variable and perform the statistical test proposed by McCrary (2008). The left graph plots the unrestricted monthly frequency and the estimated monthly frequency from a local linear regression applied to each side of the cut-off date for eligibility to temporary DI, March 2004. This graph shows that the number of recipients awarded DI changes little from September 2003 to February 2004. In particular, there is no evidence of a discontinuous change in the frequency at January 2004, as manipulation of the application date would imply: while 220 individuals were awarded DI in December 2003, 215 individuals were awarded DI in January 2004. As expected, the sample size declines after February 2004, as the introduction of temporary DI crowds out some of the use of permanent DI benefits.

uals who were awarded permanent DI after February 2004. The same holds true if we include individuals who were awarded temporary DI in the control group. The results are available upon request.

¹⁸For example, the correlation between the local DI rate and labor force participation in our sample is -0.0029 in 2005, -0.0105 in 2006, and 0.0022 in 2007.

The right graph displays the estimates behind McCrary's test. We begin by plotting the monthly density, and then smooth the histogram using local linear regression, separately on each side of the cut-off. The formal test is implemented as a Wald test of the null hypothesis that there is no jump in the density in January 2004. We cannot reject this null hypothesis: for instance, with a bandwidth of 2 months and a bin-size of 1 month, the estimated jump in the density is -0.00712 with a standard error of 0.151.

III Empirical results

A Labor force participation and earnings

The first column of Table 1 reports results from the FD model, comparing the outcomes for the recipients who were awarded DI in December 2003 (the treatment group) and January 2004 (the control group). There are substantial differences between the labor market outcomes in the two groups. The estimated effect of the return-to-work program on labor force participation steadily increased from 2-3 percentage points in 2005 to more than 5 percentage points in 2007. This pattern is consistent with standard search-theoretical models of the labor market in which it takes time for workers to find jobs and for firms to fill vacancies. By comparison, only 2-3 percent of the control group participated in the labor force, and labor force participation among the rejected applicants was approximately 30 percent. We can also see that the program almost doubled the participants' average earnings.

Insert Table 1 approximately here

The third column of Table 1 reports the RD estimates with 2 months of bandwidth on each side of the cut-off date. The key difference between the FD model and the RD model is their assumptions regarding how the potential outcomes change with the assignment variable: that is, the shape of the potential outcome curves. The FD specification takes the average outcome for the

control group as a counterfactual for the treatment group in the absence of the return-to-work program; the potential outcome curves are assumed to be flat. If the potential outcome curves instead slope upward, as the graphical evidence indicates, the FD model will underestimate the effect of the return-to-work program because the average outcome for the control group will be too large to serve as an appropriate counterfactual for the treatment group. In this type of “smoothly contaminated” randomized experiment, the RD specification uses the observed trends in the outcomes on each side of the cut-off to construct an appropriate counterfactual. As is evident from column 3 of Table 1, the RD estimates are somewhat larger than the FD estimates. Again, the estimated effect of the return-to-work program on labor force participation steadily increased: in 2005, the estimate is approximately 3 percentage points, while in 2007, it approaches 9 percentage points. We can also see that the RD model estimates the increase in earnings to be approximately 110 percent.

Columns 2 and 4 of Table 1 add a set of pre-determined covariates to the FD model and the RD model, respectively. As a consequence of a local randomized experiment, assignment to treatment near the cut-off should be independent of pre-determined covariates. Thus it is reassuring to see that our results are robust to the inclusion of the covariates. Throughout the paper, we allow for a flexible functional form by including a large set of dummies for different values of the covariates. In the remainder of the paper, we use FD and RD models with covariates as our baseline specifications: including covariates in the RD design reduces the residual variance and is thus a useful way to gain precision, especially in the subsample analysis where the sample size is relatively small (see e.g. Froelich, 2007).

B Robustness analysis

To increase the confidence in the RD design, we show that our findings are robust to a number of specification checks. We begin by estimating a difference-in-differences (DiD) specification of both the FD model and the RD model. The main motivation for this robustness check is that seasonality in applications could generate (discontinuous) changes in the potential outcomes ac-

ording to the month in which individuals were awarded DI. Although neither the award rate nor the observable characteristics of recipients show signs of seasonality in applications in the window surrounding the reform, we cannot rule out (discontinuous) changes in unobservables by month of award. The DiD specification exploits the fact that recipients who were awarded DI just before and just after January 1 of 2003 were assigned to the return-to-work program: significant differences in the post-assignment outcomes for the two groups would therefore be unrelated to the return-to-work program and should instead capture changes in outcomes due to the month in which DI was awarded.

A standard DiD estimate is obtained by estimating the difference in outcomes between individuals awarded DI in December 2002 and January 2003, and subtract it from the FD estimate of the return-to-work program. The DiD specification of the RD model is implemented in a similar way: we first re-estimate equations (1) and (2) using data for recipients awarded DI in November-December 2002 and January-February 2003, respectively; we then subtract the difference between the estimated regression intercepts on the two sides of the cut-off date from the RD estimate of the return-to-work model; and calculate the standard errors by bootstrap with 1,000 replications. Appendix Table A.2 reports estimates from the DiD specification of both the FD model and the RD model. The results suggest that month of award effects do not confound the conclusions drawn about whether financial incentives encourage DI recipients to return to work.

The DiD specification identify the impact of the return-to-work program under the assumption that month of award effects do not vary across years. Appendix Table A.3 challenges this assumption by pretending that the return-to-work program takes effect on January 1 of 2003 (instead of the true reform date, January 1 of 2004). The placebo test for the standard DiD model is performed by estimating the difference in outcomes between individuals awarded DI in December 2001 and January 2002, and subtract it from the difference in outcomes between individuals awarded DI in December 2002 and January 2003. The placebo test for the DiD specification of the RD model is similar, except that it uses the sample of individuals who were awarded DI between November

2001 and February 2002 and between November 2002 and February 2003. Both placebo tests exploit that everyone in the estimation samples were assigned to the return-to-work program: significant placebo tests are therefore a sign of month of award effects that vary across years, rather than true program impacts. It is therefore reassuring that the estimates from the placebo tests are small and insignificant.

Next, Appendix Table A.4 reports results from the usual specification checks of the RD model. Columns 1 and 2 in Panel (a) extend the bandwidth on the left side of the cut-off date by one month: the RD estimates change little, but our precision increases because the sample size (in the treatment group) increases. Column 3-6 in Panel (a) use a linear functional form of f_r and f_l . It is reassuring that the parametric regression produces estimates that are similar to those of the baseline RD specification. Columns 1-6 in Panel (b) constrain f_r and f_l to be the same on both sides of the cut-off date. Constraining the slope is in some sense inconsistent with the spirit of the RD design because the data from the right of the cutoff are used to estimate α_l , which is defined as a limit when approaching from the left of the cutoff, and vice versa. One possible justification for a common slope is that if the functional form is the same on both sides of the cutoff, then more efficient estimates of program impact are obtained by imposing that constraint. Indeed, our results are quite similar when we impose a common linear, quadratic or cubic slope.

In our analysis, we excluded individuals who (prior to DI benefit award) resided in a county with a wage-subsidy policy. The reason for this sample restriction is to avoid confounding the effects of the return-to-work program with the impact of wage subsidies to firms. Five (of nineteen) counties provided wage subsidies to firms that hired DI recipients. Subsidies were available if either the firm or the DI recipient resided in a county with the wage subsidy policy. To explore potential spillover effects between the return-to-work program and the wage subsidies, we have collected information about local labor market regions. The main criteria used for defining the local labor market regions are commuting patterns and travel times. There are 161 local labor market regions in Norway. Within five of the regions, there are counties both with and without

the wage subsidy policy. To check for potential spillovers, we re-estimate the FD and RD model without DI recipients who reside in any of these five regions. The estimates are close to our baseline results, suggesting that spillover effects are unlikely to drive the estimated effects of the return-to-work program.

Lastly, we assess to what extent the baseline RD estimates are generalizable to DI recipients with longer spells of DI at the time of the return-to-work program. As shown in Lee and Lemieux (2010), in the presence of heterogeneous treatment effects, the RD estimate can be interpreted as a weighted average treatment effect across all individuals in the estimation sample. By extending the bandwidth, the RD estimate assigns weight to the effects of the return-to-work program also for individuals “away from the threshold”. Appendix Figure A.3 shows RD estimates where we increase the sample size in the treatment group by expanding the bandwidth to the left of the threshold. We run 25 RD regressions, where each estimate increases the bandwidth to the left of the cut-off date by one month. Whether we use a bandwidth of 2 months (November and December 2003) or 26 months (November 2001 - December 2003), we always find sizeable RD estimates of the return-to-work program. The estimates decline somewhat with the size of the bandwidth, but remains substantial even with a window of 26 months; one explanation for this decline is that the costs and benefits of working may change with the length of time spent on DI, either because of human capital depreciation or due to deteriorating health. From a policy perspective, these findings complement each other: while the RD estimate with a narrow bandwidth may be more informative about the expected effects on recent or new entrants to DI, the RD estimates with wider bandwidths speak to the effects on recipients with longer spells of DI at the time of the reform.

C Implied elasticity of labor force non-participation

Table 1 demonstrated that many DI recipients have considerable capacity to work that can be effectively induced by providing financial work incentives. To compare the labor supply responses with the changes in financial incentives, we calculate the aggregate elasticity of labor force non-

participation to the participation tax rates (PTR).¹⁹

We define the aggregate elasticity of labor force non-participation as

$$(4) \quad \epsilon = \frac{\Delta(1-LFP)/(1-LFP_{control})}{\Delta PTR/PTR_{control}},$$

where $\Delta(1 - LFP)$ denotes the difference in the labor force non-participation rate between the treatment group and the control group, $(1 - LFP_{control})$ is the non-participation rate of the control group, ΔPTR denotes the difference between the treatment and the control group in terms of PTR , and $PTR_{control}$ is the mean PTR of the control group. To compute ϵ , we need to decide how to measure PTR . We begin by setting the amount of disposable income if one is not working (I_0) equal to the mean disposable income among the individuals in the control group who earn less than the SGA threshold and thus are classified as not working.²⁰ This allows us to compute a measure of PTR at a given earnings level

$$(5) \quad PTR_k = 1 - \frac{I_k - I_0}{W_k}, \quad W_k > SGA.$$

where I_k denotes the disposable income with earnings W_k .²¹ Figure 5 displays the PTR figures by earnings level. We see that the tax and benefit systems create virtually no financial reward for work in the control group. The return-to-work program lowers the PTR considerably, especially when an individual's earnings are close to the SGA threshold.

Insert Figure 5 approximately here

¹⁹In situations in which there are notches (or kinks) in the budget constraints, the PTR is generally viewed as more relevant to behavioral participation responses than the marginal tax rates (see e.g. Eissa, Kleven and Kreiner, 2006; Keane, 2011). We focus on the elasticity of labor force non-participation because it is less sensitive to small changes in the very low labor force participation rate of the control group, as compared to the labor force participation elasticity.

²⁰The elasticities barely move if we instead set I_0 equal to zero.

²¹The computation of the participation tax rate incorporates DI benefits, dependent benefits, and taxes. Both earnings and DI benefits are subject to taxation. Employers are liable for payroll tax on earnings (up to 14 percent), and there is no payroll tax on benefits. If the payroll tax is passed on to workers in the form of lower wages, then the effective tax rates are higher than what we compute.

We then define

$$(6) \quad \Delta PTR = \sum_{k > SGA} [E(PTR_k | treatment) - E(PTR_k | control)] p_k$$

where $E(PTR_k | treatment)$ is the average PTR under the return-to-work program rules with W_k in earnings, $E(PTR_k | control)$ is the average PTR under the current rules with W_k in earnings, and

$$p_k = \frac{Pr(k \leq W < k + dk | control)}{\sum_{k > SGA} Pr(k \leq W < k + dk | control)}$$

where dk is set equal to half of the SGA amount (USD 6,250 per year). This means that ΔPTR is the weighted sum of the differences between the PTRs of the treatment and control groups, with the weights given by the (conditional) density of the earnings of individuals in the control group who work. Figure 5 displays the density of earnings (above the SGA threshold) in the control group and the treatment group. We see that most of the density is quite close to the SGA threshold, which suggests that the differences in PTR in this area are weighted heavily in ΔPTR .

Insert Table 2 approximately here

Table 2 shows the elasticities implied by the estimated effects on labor force participation. Consistent with the impact on labor force participation, the implied elasticities are larger (in absolute value) for the RD model than for the FD model. The results suggest an average elasticity of labor force non-participation to PTR between 0.119 (FD model) and 0.186 (RD model) over the period 2005-2007. The elasticity estimates peak in 2007, at which time the RD model suggests that a 10 percent reduction in PTR decreases labor force non-participation by 3 percent.

D Distributional effects

So far, we have focused on the mean impacts of the return-to-work program. Since basic labor supply theory suggests that these mean impacts might average together negative and positive labor

supply responses, we turn attention to the distributional effects of the return-to-work program.²² A quantile treatment effects analysis is not feasible since few control group members are working. Instead, we explore effects across the earnings distribution by estimating the probability that earnings fall in different segments of the budget constraint (see Figure 1 and the discussion in subsection I.B).

Table 3 reports FD and RD estimates at different parts of the earnings distribution. We can see that the return-to-work program had the largest impact on the population share with earnings between C and E. In particular, more than 70 percent of the effect on labor force participation can be attributed to changes in the probability that earnings fall in this segment. This finding is consistent with the prediction that assignment to the return-to-work program increases the labor supply of DI recipients who would have located at the budget segment BC under current rules. There is also suggestive evidence of small increases in the population shares with earnings between E and e or above e. Although not statistically significant at conventional levels, these findings indicate that moves toward full-time work among individuals who would have located at segments BC or DE under current rules dominate any negative labor supply response among (the few) individuals who under current rules would have located above E.

Insert Table 3 approximately here

The graphical evidence displayed in Figure 5 complements the regression results reported in Table 3. This figure shows the density of earnings in the control group and the treatment group, alongside the PTR under current rules and in the return-to-work program. The graphical evidence suggests that the return-to-work program had the largest impact on earnings near the SGA threshold (i.e. the budget segment between C and E) where PTR decreased the most.

²²Bitler, Gelbach and Hoynes (2006) show the usefulness of estimating distributional effects in situations where theory predicts systematic heterogeneity in the impact of welfare reforms.

E Financial costs and benefits

Advocates of return-to-work policies claim that the welfare of DI recipients would improve and the cost of DI programs would decrease if DI recipients who returned to work were allowed to keep a significant fraction of their benefits. To investigate these claims, we examine the impact of the return-to-work program on disposable income, DI benefits received, taxes paid, and program costs. Table 4 shows the results. Our estimates suggest that the gains in earnings translated into a significant reduction in the cost of the program and a small and statistically insignificant increase in disposable income. The estimated reduction in costs is in the range of 3.5 and 5 percent of the costs associated with DI under current rules, and it reflects a significant decrease in benefits and a small increase in taxes paid by DI recipients.

Insert Table 4 approximately here

Critics of return-to-work policies counter that allowing DI recipients the option to return to work and keep a fraction of their benefits makes the program more generous, both for potential applicants and for current recipients; in theory, therefore, return-to-work policies could lead to fewer exits from DI by current recipients and encourage more applications for DI. Note, however, that there is little room for a reduction in the exit rate, as few DI recipients exited the DI program under current rules. Indeed, we find no evidence of any impact of the return-to-work program on the rate of exit from DI during the period 2005-2008. The point estimates are small and positive (0.01 in FD model and 0.03 in RD model), but too imprecisely estimated to draw firm conclusions. Unfortunately, our study is not informative about the level of induced entry that might occur if new DI recipients were made eligible for the return-to-work program. In Appendix B, however, we make a calculation for the size of the induced entry that would have to be generated by the return-to-work policy to lead to an increase in program costs. Our calculation suggests that the elasticity of entry response to the change in DI benefits needs to be much higher than what the

existing literature suggests.²³

F Heterogeneity

Table 2 demonstrated that many DI recipients are elastic to changes in financial incentives. However, these findings should be interpreted with caution because they pertain only to DI recipients who are between the ages of 18 and 49. We now investigate the impact of the return-to-work program on older DI recipients. Furthermore, we examine the degree of heterogeneity of the responses to the return-to-work program among younger DI recipients.

Appendix Table A.5 reports the regression results from the subsample analysis. As is evident from columns 3 and 4 of panel (a), much would be missed if we averaged together the labor supply responses of older and younger DI recipients. Among older DI recipients, there is no evidence of an increase in labor force participation due to the return-to-work program. One possible explanation for this is that the gains from returning to the labor force may be greater for younger DI recipients.²⁴ Another possible explanation is that the cost of working may be lower for younger DI recipients, who more often tend to suffer from difficult-to-verify disorders such as mental illness (Autor and Duggan, 2006; von Wachter, Song and Manchester, 2011). To shed some light on these explanation, we divide the sample of younger DI recipients into two groups: individuals with mental illness (40 percent) and those suffering from other disorders (60 percent).²⁵ The estimates reported in columns 5-8 of panel (b) lend support to the first explanation, as the impact of the

²³The baseline RD (FD) estimate implies a lower bound of 1.5 (1.1) on the elasticity of induced entry that would lead the return-to-work policy to increase program costs if new DI recipients were made eligible. In comparison, Gruber (2000) reports elasticity estimates in the range of 0.28-0.36, whereas Campolieti (2004) finds little, if any, effect on entry rates from an increase in benefit levels. There is also a sizeable literature that has estimated elasticities of entry response by modeling labor force participation or DI recipiency as a function of potential DI benefit levels. Using cross-sectional variation in DI benefits, these studies tend to find elasticities in the range of 0.1-0.2 (see Gruber (2000) for a review of this evidence).

²⁴Higher gains from working for younger DI recipients may either come from shorter length of time spent on DI and thus less human capital depreciation or because they have more years to potentially work and earn retirement benefits (Bound, Burkhauser and Nichols, 2003; von Wachter, Song and Manchester, 2011).

²⁵Among individuals with other disorders, the main diagnosis groups are musculoskeletal diseases (37 percent), neurological disorders (17 percent) and circulatory system disorders (5 percent). Because of small sample sizes, it is not feasible to explore heterogeneity in the effects by these disorders.

return-to-work program tends to be smaller among DI recipients suffering from mental illness.

We also explore several dimensions of heterogeneity within the group of younger DI recipients. Columns 5-8 of panel (d) show that the return-to-work program had the largest impact on DI recipients living in areas with low unemployment rates, suggesting an important link between demand side factors and program incentives. This finding is consistent with existing evidence showing that local labor market conditions have a significant impact on welfare spells for most demographic groups (see e.g. Hoynes, 2000). Previous research has also found significant differences in the responsiveness of different groups of welfare recipients to changes in financial incentives (see e.g. Moffitt, 1992). In particular, supply-side factors such as education and labor market experience tend to be important. In line with this evidence, the responses to the return-to-work program are stronger for individuals with high education levels or more labor market experience (measured by AIE or years of employment).²⁶ We further examine gender differences in the effects of the return-to-work program. Columns 1-4 of panel (b) demonstrate that males respond considerably stronger to the financial incentives than females. The relatively weak response to the return-to-work program among women is worrisome given that the growth of disability receipt in recent decades has been far more pronounced among women than men (Autor and Duggan, 2006).

Insert Table 5 approximately here

Lastly, Table 5 summarizes the heterogeneity in program effects: For each subgroup, it displays the average effect on labor force participation (ΔLFP), the average change in participation tax rate (ΔPTR), and the average elasticity of labor force non-participation (ϵ), over the period 2005-2007. Table 5 shows that the heterogeneity in program effects is weakly related to differences in PTR across subgroups. Instead, we find that some subgroups respond much stronger to similar changes

²⁶Unfortunately, we do not have information on other work-related factors such as previous occupation and industry. Because of small sample sizes, we are unable to examine the effects of supply-side factors in different local labor markets. The differential effects by supply-side factors could, in theory, therefore reflect that individuals with low education levels and little labor market experience may be more likely to live in areas with adverse labor markets. However, the correlations between the local unemployment rate and individuals' labor market experience and education are weak (-0.05 for AIE and -0.04 for years of schooling).

in financial incentives; for example, women and individuals with low educational levels experience relatively large reductions in participation tax rates, but still increase their labor force participation less than males and highly educated individuals. The results in Table 5 also illustrate the problem of using cross-sectional variation in participation tax rates across groups to estimate the labor supply elasticity of DI recipients.

G Comparison of upper bound estimates

We conclude our empirical analysis with a comparison between the upper bound estimates of the counterfactual labor force participation rates of DI recipients in Norway and the U.S. This comparison puts the magnitude of the RD estimates into perspective, and sheds light on the external validity of the evidence from Norway.

We employ the approach suggested by Bound (1989), who uses a sample of rejected DI applicants as a control group for the DI recipients and considers their labor force participation rate as an upper bound estimate of the counterfactual labor force participation rate of DI recipients. On the basis of data collected in 1972 and 1978, Bound finds that the labor force participation rate of DI recipients in the U.S. would have been at most 30 percentage points higher had they not received disability benefits. Recent studies have extended this analysis, in part by analyzing different subgroups and different time periods but also with quasi-experimental variation in the disability determination process to improve the comparability of DI recipients and rejected DI applicants.²⁷

Although there is some heterogeneity in the U.S. evidence, the estimates are consistently higher than the labor force participation rates for our treatment group. There are several possible explanations. One is that rejected applicants are likely to be in better health; thus, their labor force attachment constitutes an upper bound for the employment behavior of new recipients. Another possible explanation is that the counterfactuals differ: in Bound's approach, the counterfactual

²⁷See Chen and van der Klaauw (2008), French and Song (2009), von Wachter, Song and Manchester (2011), and Maestas, Mullen and Strand (2012).

is a scenario in which the DI program does not exist, whereas in our case, the counterfactual is a scenario in which DI recipients are faced with higher participation tax rates. A third possible explanation is that DI recipients in the U.S. and Norway have substantially different work capacity and labor supply elasticity. If the latter explanation was empirically important, it would raise concerns about the external validity of our RD estimates.

To shed light on this issue of external validity, we apply the Bound approach to the Norwegian data. Appendix Table A.6 displays the mean differences between the rejected applicants and those who were awarded DI in terms of observable characteristics and labor force participation. The sample contains three groups: individuals whose DI applications were rejected in December 2003 or January 2004, individuals who were awarded DI in December of 2003 (treatment group), and individuals who were awarded DI in January 2004 (control group). When we compare the rejected applicants with the control (treatment) group, the counterfactual is the scenario in which DI under the current rules (the return-to-work program rules) does not exist. A concern with the Bound approach is that the employment rate of rejected applicants might in part reflect the fact that some of these individuals were still in the process of appealing the denial (see e.g. Bound, 1991; Parsons, 1991). Following von Wachter, Song and Manchester (2011), we address this issue by excluding rejected applicants who were subsequently awarded DI benefits over the next five years.²⁸

Appendix Table A.6 shows that the labor force participation rate is 22-26 (24-30) percentage points higher among rejected applicants than among treatment group (control group). In the Bound approach, these differences are interpreted as upper bound estimates of the counterfactual labor force participation rate among DI recipients. However, there are some significant differences in the observable characteristics of the rejected and awarded DI applicants. When we control for the differences in their observable characteristics, the upper bound estimates increase by 2-4 per-

²⁸When comparing the Norwegian upper bound estimates to the U.S. evidence, it is therefore reassuring to find that appeals play a similar role in the two countries. In the U.S., 60 percent of the initially rejected applicants appeal; of these, 70 percent were subsequently awarded DI benefits over the next five years (Autor et al., 2011). In Norway, 57 percent of the initially rejected applicants appealed (or quickly re-applied); of these, 77 percent were subsequently awarded DI benefits over the next five years.

centage points. These estimates are quite similar to the recent US evidence, and very close to the results for DI recipients of comparable ages reported in von Wachter, Song and Manchester (2011).

The similarity between the upper bound estimates indicates that the work capacity and labor supply elasticity of DI recipients in Norway are comparable to those of DI recipients in the U.S., which lends support to the external validity of our analysis of the return-to-work program. This suggests that our findings may have important implications for the proposed “\$1 for \$2 offset” policy in the U.S. As shown in Appendix Figure A.1, the U.S. policy proposal strengthens the incentives for labor force participation more than the return-to-work program. Our findings therefore suggest that the “\$1 for \$2 offset” policy may be quite effective in inducing DI recipients in the U.S. to return to work.

IV Conclusion

Autor and Duggan (2006) argue that the liberalization of the screening process and the movement away from physically exertive work have blurred the divide between the totally and permanently disabled and those who are disabled but retain some work capacity. In this paper, we have investigated whether many individuals who are deemed totally and permanently disabled by the Social Security Administration indeed have considerable work capacity and examined how elastic their labor supply is to changes in financial incentives.

Using a local randomized experiment that arises from a sharp discontinuity in DI policy in Norway, we provide transparent and credible identification of how financial incentives induce DI recipients to return to work. We find that many DI recipients have considerable capacity to work that can be effectively induced by providing financial work incentives. We also show that encouraging DI recipients to work not only increases their disposable income but also reduces program costs.

However, these findings should be interpreted with caution: they apply only to DI recipients

between the ages of 18 and 49. Among older DI recipients, we find no response to financial work incentives. However, because the age of entry into DI has decreased in most developed countries during the last few decades, the elastic labor supply of younger DI recipients may be quite important from a policy perspective. In particular, our findings indicate that policies that target younger DI recipients may be most effective in encouraging DI recipients to return to work.

Another caveat is that our study is not informative about the level of induced entry that might occur if DI recipients are given the option to return to work and keep a fraction of their benefits. This is because the cut-off date for eligibility to the return-to-work program was retroactive. We do, however, make a calculation for the size of the induced entry that would have to be generated by the return-to-work policy to lead to an increase in program costs. Our calculation indicates that the elasticity of entry response to the change in DI benefits needs to be much higher than what the existing literature suggests.

Lastly, we advise readers to exercise the usual caution in importing our findings to other countries. Nevertheless, as shown above, the DI programs in Norway and the U.S. are similar in many respects. Furthermore, when we apply the Bound (1989) approach to the Norwegian data, we find upper bound estimates of the counterfactual labor force participation rate of DI recipients that are fairly consistent with recent U.S. evidence. This finding indicates that the work capacity and labor supply elasticity of DI recipients in Norway are comparable to those of DI recipients in the U.S., which lends some support to the external validity of our analysis of the return-to-work program.

References

Adam, Stuart, Antoine Bozio, and Carl Emmerson. 2011. "Reforming Disability Insurance in the UK: Evaluation of the Pathways to Work Programme." IFS Working Paper.

- Autor, David.** 2011. “The Unsustainable Rise of the Disability Rolls in the United States: Causes, Consequences, and Policy Options.” NBER Working Paper 17697.
- 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): 157–205.
- Autor, David H., and Mark G. Duggan.** 2006. “The growth in the social security disability rolls: A fiscal crisis unfolding.” *Journal of Economic Perspectives*, 20(3): 71–96.
- Autor, David, Nicole Maestas, Kathleen Mullen, and Alexander Strand.** 2011. “Does Delay Cause Decay? The Effect of Administrative Decision Time on the Labor Force Participation and Earnings of Disability Applicants.” University of Michigan Retirement Research Center Working Paper 2011-258.
- Benitez-Silva, Hugo, Moshe Buchinsky, and John Rust.** 2010. “Induced Entry Effects of a \$1 for \$2 Offset in SSDI Benefits.” SUNY-Stony Brook Working Paper.
- Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes.** 2006. “What mean impacts miss: Distributional effects of welfare reform experiments.” *American Economic Review*, 96(4): 988–1012.
- Black, Dan, Kermit Daniel, and Seth Sanders.** 2002. “The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust.” *American Economic Review*, 92(1): 27–50.
- Borghans, Lex, Anne Gielen, and Erzo Luttmer.** 2012. “Social support substitution and the earnings rebound: Evidence from a regression discontinuity in disability insurance reform.” NBER Working Paper 18261.
- Bound, John.** 1989. “The Health and Earnings of Rejected Disability Insurance Applicants.” *American Economic Review*, 79(3): 482–503.

- Bound, John.** 1991. “The Health and Earnings of Rejected Disability Insurance Applicants - Reply.” *American Economic Review*, 81(5): 1427–1434.
- Bound, John., Richard Burkhauser, and Austin Nichols.** 2003. “Tracking the Household Income of SSDI and SSI Applicants.” In *Research in Labor Economics*. Vol. 22, , ed. Solomon W. Polachek, 113–158. Emerald Group Publishing.
- Bratberg, Espen.** 1999. “Disability retirement in a welfare state.” *Scandinavian Journal of Economics*, 101(1): 97–114.
- Bratsberg, Bernt, Elisabeth Fevang, and Knut Roed.** 2010. “Disability in the Welfare State: An Unemployment Problem in Disguise?” IZA Discussion Paper 4897.
- Campolieti, Michele.** 2004. “Disability insurance benefits and labor supply: Some additional evidence.” *Journal of Labor Economics*, 22(4): 863–889.
- 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: 306–316.
- 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): 757–784.
- Eissa, Nada, Henrik J. Kleven, and Claus T. Kreiner.** 2006. “Welfare Effects of Tax Reform and Labor Supply at the Intensive and Extensive Margins.” In *Tax policy and labor market performance*. , ed. Jonas Agell and Peter Birch Sørensen. MIT Press.
- French, Eric, and Jae Song.** 2009. “The Effect of Disability Insurance Receipt on Labor Supply.” FRB of Chicago Working Paper 2009-05.
- Froelich, Markus.** 2007. “Regression Discontinuity Design with Covariates.” University of St. Gallen Department of Economics Working Paper 2007-32, SSRN.

- Fuji, Daisuke, Guido Imbens, and Karthik Kalyanaraman.** 2009. “Notes for Matlab and Stata Regression Discontinuity Software.” Harvard University.
- Gruber, Jonathan.** 2000. “Disability insurance benefits and labor supply.” *Journal of Political Economy*, 108(6): 1162–1183.
- Hoynes, Hilary W., and Robert Moffitt.** 1999. “Tax rates and work incentives in the Social Security Disability Insurance Program: Current law and alternative reforms.” *National Tax Journal*, 52(4): 623–654.
- Hoynes, Hilary Williamson.** 2000. “Local Labor Markets and Welfare Spells: Do Demand Conditions Matter?” *Review of Economics and Statistics*, 82(3): 351–368.
- Imbens, Guido, and Karthik Kalyanaraman.** 2012. “Optimal Bandwidth Choice for the Regression Discontinuity Estimator.” *The Review of Economic Studies*, 79(3): 933–959.
- Keane, Michael P.** 2011. “Labor Supply and Taxes: A Survey.” *Journal of Economic Literature*, 49(4): 961–1075.
- Lee, David S., and Thomas Lemieux.** 2010. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature*, 48(2): 281–355.
- Livermore, Gina, Allison Roche, and Sarah Prenovitz.** 2009. “SSI and DI Beneficiaries with work related goals and expectations.” In *Work activity and use of employment supports under the original Ticket to Work Regulations*. Mathematica Policy Research.
- Maestas, Nicole., Kathleen Mullen, and Alexander Strand.** 2012. “Does Disability Insurance Discourage Work?” *American Economic Review*, Forthcoming.
- McCrary, Justin.** 2008. “Manipulation of the running variable in the regression discontinuity design: A density test.” *Journal of Econometrics*, 142: 698–714.

- Milligan, Kevin S., and David A. Wise.** 2011. "Social Security and Retirement around the World: Mortality, Health, Employment and Disability Insurance Participation and Reforms." Harvard Kennedy School Working paper.
- Moffitt, Robert.** 1992. "Incentive Effects of the U.S. Welfare System: A Review." *Journal of Economic Literature*, 30(1): 1–61.
- Moore, Timothy J.** 2011. "The Employment Effects of Terminating Disability Benefits: Insights from Removing Drug Addictions as Disabling Conditions." University of Maryland Working Paper.
- NOU.** 2007. "Ny uførestønning og ny alderspensjon til uføre." White Paper NOU 2007:4 (In Norwegian).
- Parsons, Donald O.** 1991. "The Health and Earnings of Rejected Disability Insurance Applicants: Comment." *American Economic Review*, 81(5): 1419–1426.
- Rege, Mari, Kjetil Telle, and Mark Votruba.** 2009. "The Effect of Plant Downsizing On Disability Pension Utilization." *Journal of the European Economic Association*, 7(4): 754–785.
- Stapleton, David, Gina Livermore, Craig Thornton, Bonnie O’Day, Robert Weathers, Krista Harrison, So O’Neil, Emily S. Martin, David Wittenburg, and Debra Wright.** 2008. "Ticket to Work at the Crossroads: A solid foundation with an uncertain future." Report delivered to the Social Security Administration, Washington DC. Matematica Policy Research.
- 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.

Table 1: PROGRAM IMPACT ON LABOR FORCE PARTICIPATION AND EARNINGS

<i>Outcome variables:</i>	FD		RD		Comparison means [st. dev]	
	FD	FD w/c	RD	RD w/c	Jan. & Feb. 04	Rejected applicants
	LFP(2005)	0.022 (0.015)	0.033** (0.016)	0.028 (0.024)	0.038 (0.025)	0.018 [0.134]
LFP(2006)	0.031* (0.017)	0.033* (0.018)	0.039 (0.027)	0.042* (0.026)	0.02 [0.142]	0.316 [0.466]
LFP(2007)	0.054*** (0.02)	0.053** (0.022)	0.087*** (0.031)	0.085*** (0.031)	0.034 [0.182]	0.316 [0.466]
Avg. earnings (2005-2007)	1126** (508)	1247** (535)	1630** (799)	1644** (781)	1551 [5033]	13223 [21314]
<i>Characteristics:</i>	FD		RD		Comparison means	
	FD	FD w/c	RD	RD w/c	Jan. & Feb. 04	Rejected applicants
	Age at DI award	-0.30 (0.91)	-0.20 (1.44)	38.4 [9.7]	38.6 [7.5]	
Male	-0.052 (0.048)	-0.076 (0.075)	0.50 [0.50]	0.515 [0.50]		
Years of Schooling	0.32 (0.28)	0.29 (0.44)	10.5 [3.1]	9.9 [3.8]		
Experience	-0.21 (0.92)	-0.43 (1.45)	13.9 [10.0]	11.3 [9.0]		
AIE	-290 (1565)	-1123 (2486)	38013 [17686]	34558 [15118]		
Local unemployment rate	-0.001 (0.001)	-0.002 (0.001)	0.024 [0.008]	0.024 [0.009]		
Local DI rate	0.004 (0.003)	0.005 (0.004)	0.098 [0.024]	0.098 [0.024]		
Number of Children	-0.031 (0.11)	-0.08 (0.17)	0.90 [1.13]	0.954 [1.23]		
Musculoskeletal system	0.013 (0.040)	0.026 (0.062)	0.22 [0.415]	0.332 [0.472]		
Mental disorders	0.041 (0.048)	0.041 (0.075)	0.386 [0.487]	0.321 [0.468]		
Obs	435		897		440	196

*** significant at 1% level, ** significant at 5% level, *significant at 10% level.
Standard errors (in parentheses) are robust to heteroscedasticity.

Note: This table displays the main results. The first two columns display estimates from the first difference model without covariates (FD) and with covariates (FD w/c); the sample consists of individuals aged 18-49 (at the date of DI award), who were awarded DI in Dec. 2003 & Jan. 2004. The third and fourth columns display estimates from the regression discontinuity model without covariates (RD) and with covariates (RD w/c); we use a local linear regression with a triangular kernel density function and 2 months of bandwidth on each side of the cut-off date; the sample consists of individuals aged 18-49 (at the date of DI award), who were awarded DI in Nov. 2003 - Feb. 2004. Labor force participation (LFP) is equal to one if annual earnings exceed SGA. Average earnings are the mean annual earnings in USD over the period 2005-2007 (NOK/USD ≈6). All individual characteristics are measured prior to disability award. Experience is the number of years with earnings above SGA. The AIE summarizes the earning history. The local unemployment rate is the fraction of the adult population (aged 18-66) receiving unemployment benefits in the municipality of residence. The local DI rate is the fraction of the eligible population receiving DI benefits in the municipality of residence. Number of children is the total number of children under 18. We also report the proportion of DI recipients with the two main diagnosis groups: musculoskeletal diseases and mental disorders. We allow for a flexible functional form by including a large set of dummies for different values of the covariates. The fifth column shows mean and standard deviation (in square brackets) of the outcomes and characteristics for individuals aged 18-49 (at the date of DI award), who were awarded DI in Jan. & Feb. 2004. The last column shows mean and standard deviation (in square brackets) of the outcomes and characteristics for individuals aged 18-49 (at the date of DI rejection), whose DI application was rejected in Dec. 2003 & Jan. 2004 (excluding those who successfully re-apply/appeal during the next five years).

Table 2: IMPLIED ELASTICITY OF LABOR-FORCE NON-PARTICIPATION

	FD model				RD model			
	2005	2006	2007	Average	2005	2006	2007	Average
$\Delta(1 - LFP)$	-0.033	-0.033	-0.053	-0.040	-0.038	-0.042	-0.085	-0.055
$(1 - LFP_{control})$	0.986	0.981	0.981	0.983	0.981	0.979	0.966	0.976
ΔPTR	-0.251	-0.294	-0.332	-0.292	-0.244	-0.285	-0.250	-0.260
$PTR_{control}$	0.847	0.857	0.865	0.855	0.849	0.857	0.863	0.856
Elasticity(ϵ)	0.114	0.099	0.141	0.119	0.135	0.130	0.300	0.186

Note: This table displays aggregate non-participation elasticities for the sample of individuals aged 18-49 at the date of DI award. The elasticity formula is given in equation (4). The differences in labor force non-participation between the treatment and the control group ($\Delta(1 - LFP)$) are from the FD and RD model with covariates (see the second and fourth columns of Table 1). The participation tax rate (PTR) is given by equation (5), and incorporates the DI system, income taxation, and dependent benefits. The differences in PTR between the treatment and control group (ΔPTR) are given by equation (6). $(1 - LFP_{control})$ denotes the non-participation rate of the control group, while $PTR_{control}$ is the mean PTR of the control group.

Table 3: DISTRIBUTIONAL EFFECTS

FD model				
	<i>Earnings > C</i>	<i>C < Earnings ≤ E</i>	<i>E < Earnings ≤ e</i>	<i>Earnings > e</i>
Earnings(2005)	0.033** (0.016)	0.026 (0.016)	0.011 (0.012)	0 (0.000)
Earnings(2006)	0.033* (0.018)	0.025 (0.017)	0.007 (0.005)	0.009 (0.009)
Earnings(2007)	0.053*** (0.022)	0.038** (0.019)	0.012* (0.007)	0.015 (0.011)
Avg. earnings (2005-2007)	0.062*** (0.019)	0.053*** (0.018)	0.016 (0.011)	0.006 (0.006)
Obs	435	435	435	435
RD model				
	<i>Earnings > C</i>	<i>C < Earnings ≤ E</i>	<i>E < Earnings ≤ e</i>	<i>Earnings > e</i>
Earnings(2005)	0.038 (0.025)	0.028 (0.024)	0.009 (0.018)	0.01 (0.008)
Earnings(2006)	0.042* (0.026)	0.031 (0.023)	0.01 (0.008)	0.011 (0.011)
Earnings(2007)	0.085*** (0.031)	0.062** (0.027)	0.019* (0.010)	0.022 (0.016)
Avg. earnings (2005-2007)	0.089*** (0.028)	0.074*** (0.025)	0.024* (0.014)	0.01 (0.009)
Obs	897	897	897	897

*** significant at 1% level, ** significant at 5% level, *significant at 10% level.
Standard errors (in parentheses) are robust to heteroscedasticity.

Note: This table displays the estimated impact of the return-to-work program on the probability that earnings fall in different segments of the budget constraint (see figure 1) over the period 2005-2007. The sample consists of individuals aged 18-49 at the date of DI award. The estimates are from the FD and RD model with covariates (see Table 1).

Table 4: FINANCIAL COSTS AND BENEFITS

<i>Outcome variables:</i>	Estimates		Comparison means [st. dev]	
	FD-model	RD-model	Jan. & Feb 04	Rejected Applicants
Avg. Earnings	1247**	1644**	1551	13223
(2005-2007)	(535)	(782)	[5032]	[21314]
Avg. Benefits	-798**	-1155**	33770	N/A
(2005-2007)	(359)	(536)	[8080]	
Avg. Tax	201	263	5234	4915
(2005-2007)	(156)	(229)	[2790]	[5807]
Avg. Disposable Income	248	226	30087	24402
(2005-2007)	(279)	(422)	[6217]	[19606]
Avg. Program Cost	-1000**	-1418**	28535	N/A
	(396)	(580)	[6169]	
Obs	435	897	440	196

*** significant at 1% level, ** significant at 5% level, *significant at 10% level.
Standard errors (in parentheses) are robust to heteroscedasticity.

Note: This table displays the estimated impact of the return-to-work program on average earnings, benefits, taxes, disposable income and program costs (in USD) over the period 2005-2007 (NOK/USD ≈6). The sample consists of individuals aged 18-49 at the date of DI award. The estimates are from the FD and RD model with covariates (see Table 1). Taxes include tax on earnings and benefits. Program cost equals benefits minus taxes. Disposable income equals earnings plus benefits less taxes. The third column shows mean and standard deviation (in square brackets) of the outcomes for individuals aged 18-49 (at the date of DI award), who were awarded DI in Jan. & Feb. 2004. The last column shows mean and standard deviation (in square brackets) of the outcomes for individuals aged 18-49 (at the date of DI rejection), whose DI application was rejected in Dec. 2003 and Jan. 2004 (excluding those who successfully re-apply/appeal during the following five years).

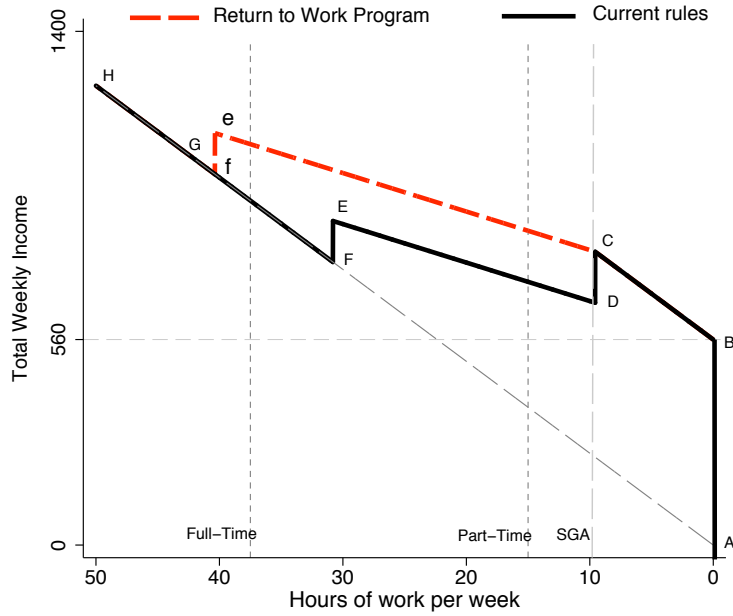
Table 5: HETEROGENEITY IN PROGRAM EFFECTS

Subgroups:	FD model			RD model		
	ΔLFP	ΔPTR	Elasticity(ϵ)	ΔLFP	ΔPTR	Elasticity(ϵ)
Females	0.017	-0.321	0.046	0.026	-0.291	0.080
Males	0.060	-0.270	0.193	0.079	-0.257	0.268
Low education	0.037	-0.297	0.110	0.041	-0.28	0.131
High education	0.061	-0.279	0.188	0.085	-0.256	0.291
Low unempl. rate	0.067	-0.285	0.204	0.111	-0.249	0.392
High unempl. rate	0.010	-0.301	0.030	0.008	-0.291	0.024
Low experience	-0.001	-0.264	0.002	0.010	-0.279	0.031
High experience	0.061	-0.317	0.170	0.079	-0.265	0.269
Low AIE	0.016	-0.287	0.049	0.028	-0.219	0.11
High AIE	0.067	-0.298	0.194	0.093	-0.281	0.293
Mental disorders	0.033	-0.313	0.091	0.039	-0.296	0.111
Other disorders	0.054	-0.265	0.176	0.067	-0.255	0.234

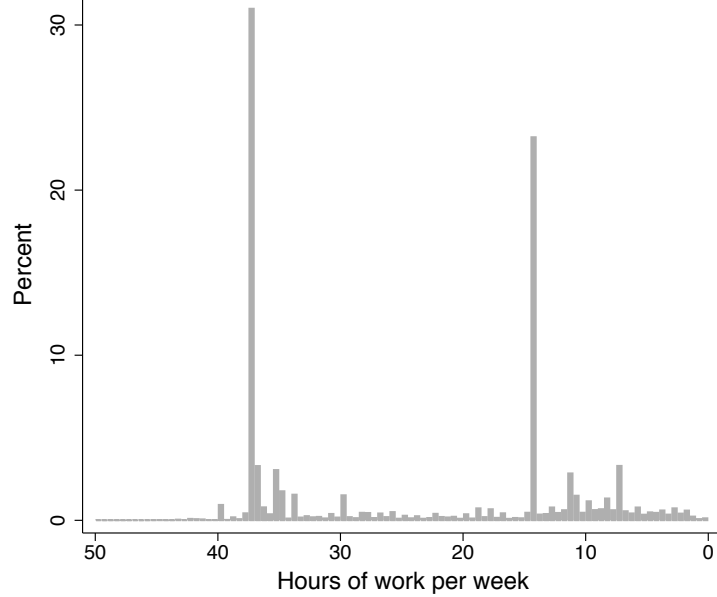
Note: This table summarizes the subsample estimates reported in Appendix Table A.5. The sample consists of individuals aged 18-49 at the date of DI award. Columns 1 and 3 report the average effect on labor force participation (ΔLFP), columns 2 and 4 report the average change in participation tax rate (ΔPTR), and columns 3 and 6 report the average elasticity of labor force non-participation (ϵ), over the period 2005-2007. The elasticity formula is given in equation (4). Columns 1-3 use the baseline FD model, while columns 4-6 use the baseline RD model. The elasticity and its components are computed separately for each subgroup (see Table 2 for a description of how the elasticity is computed). Low (high) experience is defined as years of experience below (above) the sample median. Low (high) AIE is defined as average indexed earnings below (above) the sample median. Low (high) unemployment rate is defined as living in a municipality in which the unemployment rate is below (above) the sample median. High (low) education is defined as (not) having completed high school.

Figure 1: WORK INCENTIVES

(a) Budget constraints



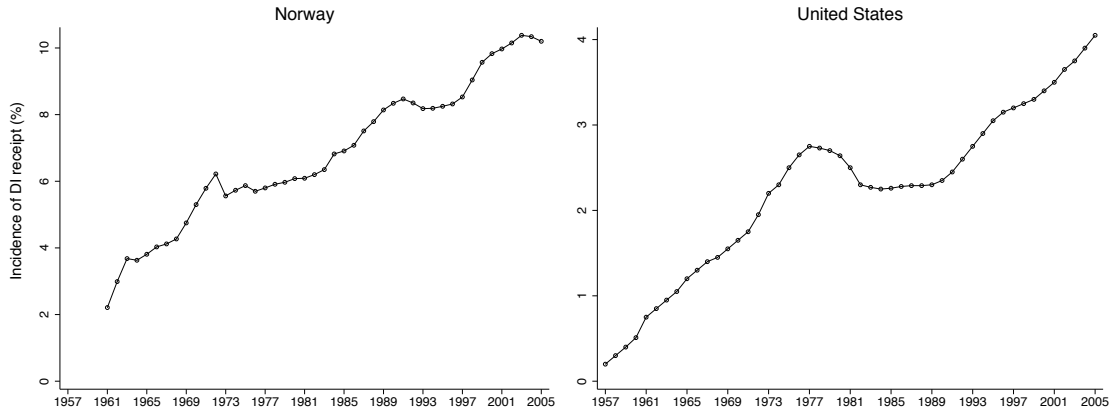
(b) Hours of work distribution of employees (not on DI)



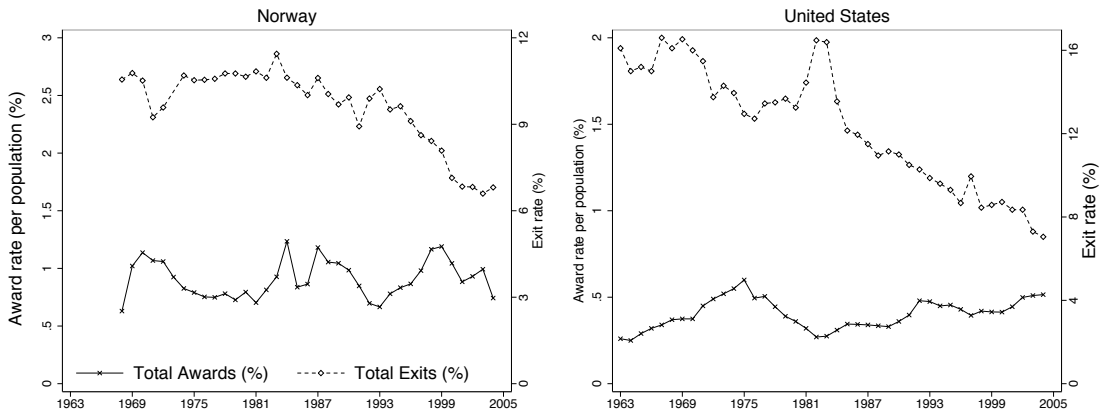
Note: GRAPH (A): The black solid lines represent the budget set for individuals under current rules and the dashed lines represent the change in the budget constraint due to the return-to-work program. We use the wage and benefit levels associated with a typical DI recipient. To compute the benefit levels, we use the average work history of recipients awarded DI during the period July 2003 - Feb. 2004. We further set the hourly wage equal to the average hourly wage in a sample of workers who are not on DI but have similar observable characteristics as the DI recipients (see footnote 8). For simplicity, we disregard income taxation and dependent benefits: Total weekly income is thus equal to earnings plus DI benefits. GRAPH (B): Displays the hours of work distribution of employees who are not receiving DI benefits. We use information from the Wage Statistics Survey in 2005. This survey covers all employees in the public sector. For employees in the private sector, the data is based on an annual stratified random sampling of all enterprises. Hours of work represent contractual number of working hours per week, excluding meal breaks.

Figure 2: DI TRENDS IN NORWAY AND THE U.S.

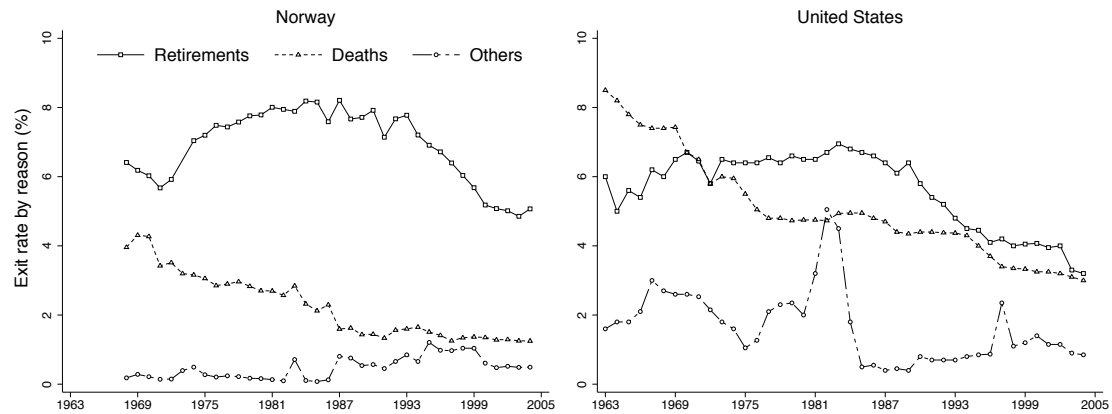
(a) Incidence of DI receipt



(b) Award and Exit Rates



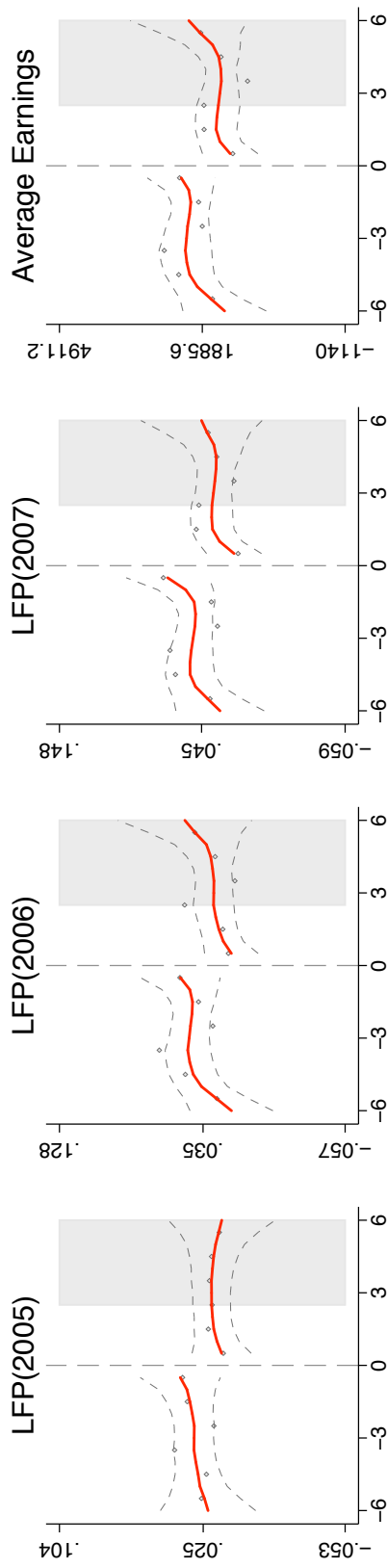
(c) Exit rates by reason



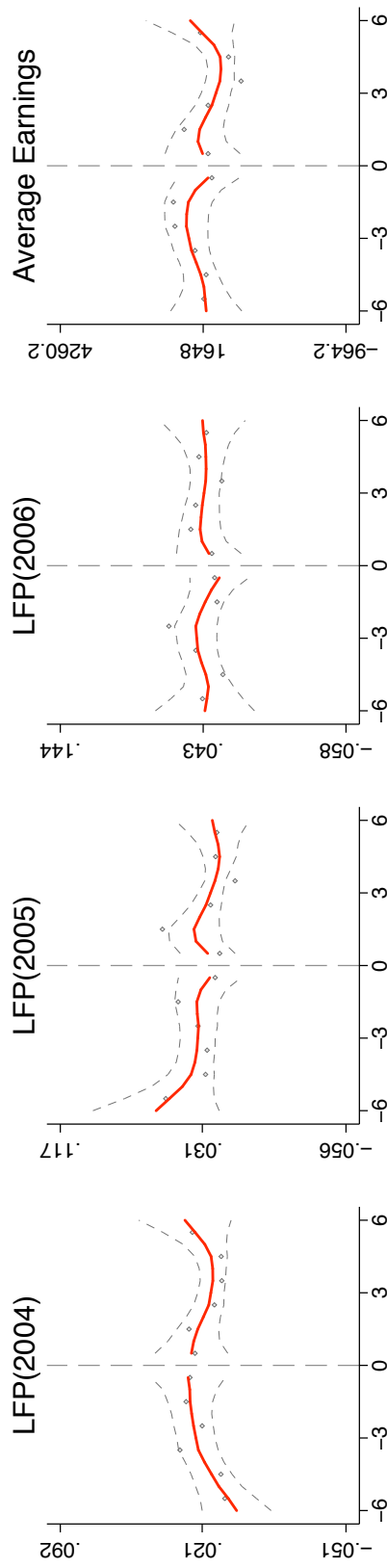
Note: The U.S. SSDI trends are based on Autor and Duggan (2006), while the Norwegian trends are collected from various issues of the Social Security Annual Statistical Supplement. Unlike the Norwegian DI program, the U.S. SSDI program does not offer temporary disability benefits. To improve comparability, Figure 2 therefore excludes temporarily disabled individuals when computing the number of DI recipients in Norway. Panel (a): Incidence of DI recipiency, defined as the percent receiving DI benefits in the adult population (aged 18 - 67 years in Norway; aged 25-64 years in the US). Panel (b): Award rate in the adult population (left axis), and exit rate from the population of DI recipients (right axis). Panel (c): Exit rates because of death, retirement, or other reasons (including eligibility-based exits).

Figure 3: LABOR MARKET OUTCOMES BY AWARD DATE

July 2003 – June 2004

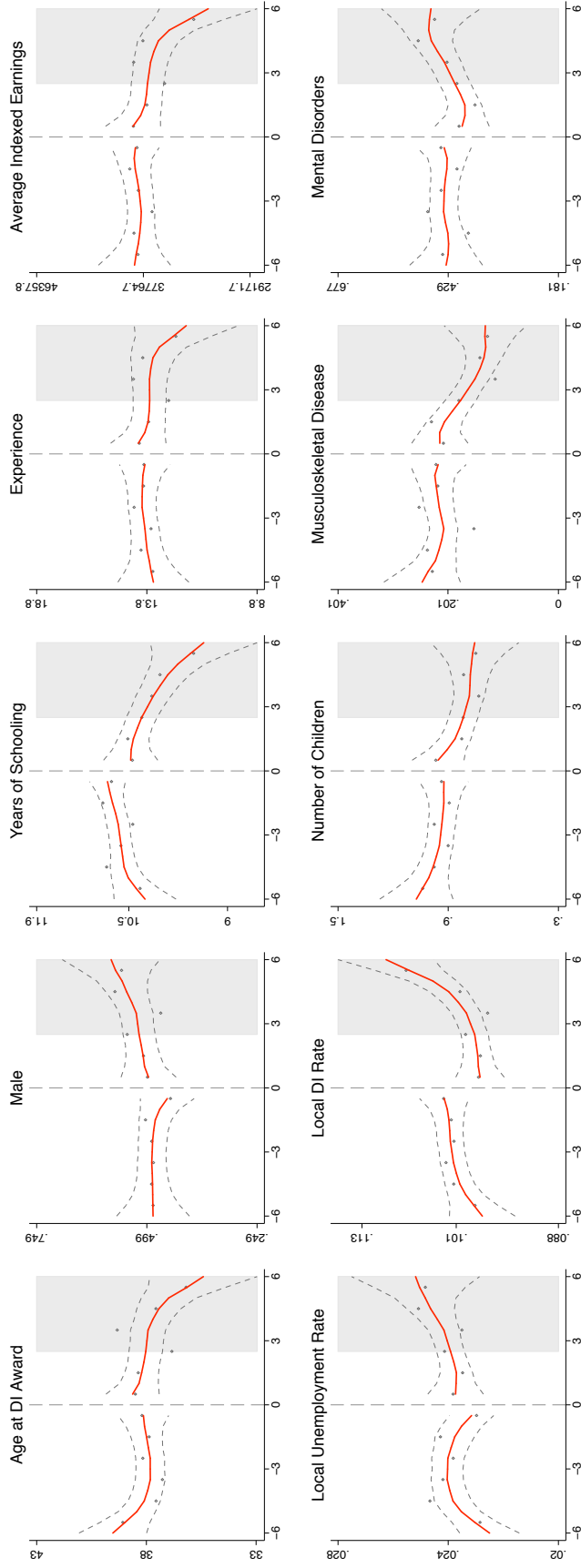


July 2002 – June 2003



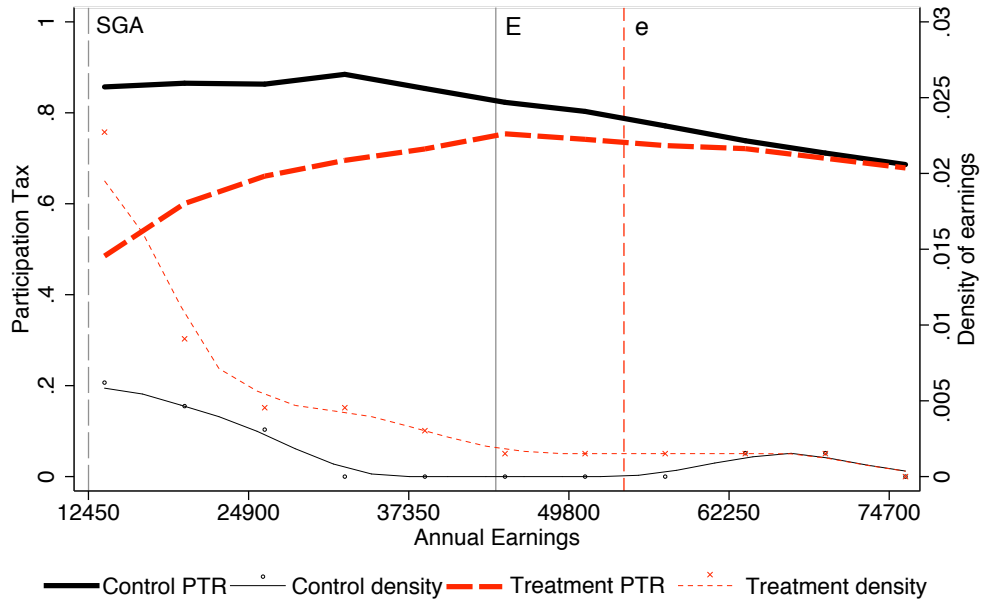
Note: The figure displays labor force participation (LFP) and mean earnings by month of DI award, six months before and after the cut-off date (denoted 0). LFP is equal to one if annual earnings exceed the SGA threshold. Average earnings (in USD) is the mean annual earnings over the period 2005-2007. The top panel consist of individuals aged 18-49 (at the date of DI award), who were awarded DI benefits during the period July 2003 to June 2004. The bottom panel consist of individuals aged 18-49 (at the date of DI award), who were awarded DI benefits during the period July 2002 to June 2003. In each graph, we plot the unrestricted monthly means and the estimated monthly means from a local linear regression applied to each side of the cut-off date (denoted 0; Jan. 1, 2004 in top panel; Jan. 1, 2003 in bottom panel). The dashed lines represent the 95 percent confidence interval of the estimated monthly means. The gray shaded areas mark the months in which individuals awarded DI were affected by a potentially confounding policy change. In each graph, the scale of the y-axis is set equal to ± 0.5 standard deviation of the respective variable.

Figure 4: CHARACTERISTICS BY AWARD DATE
 July 2003 – June 2004



Note: The figure displays individual characteristics by month of DI award, six months before and after the cut-off date, Jan. 1, 2004 (denoted 0). All individual characteristics are measured prior to disability award, and described in Table 1. The sample consists of individuals aged 18–49 (at the date of DI award), who were awarded DI benefits during the period July 2003 to June 2004. In each graph, we plot the unrestricted monthly means and the estimated monthly means from a local linear regression applied to each side of the cut-off date. The dashed lines represent the 95 percent confidence interval of the estimated regression lines. The gray shaded areas mark the months in which individuals awarded DI were affected by a potentially confounding policy change. In each graph, the scale of the y-axis is set equal to ± 5 standard deviation of the respective variable.

Figure 5: DISTRIBUTION OF EARNINGS AND PARTICIPATION TAX RATES



Note: This figure displays the the density of annual earnings (right axis) and the participation tax rate (left axis) for earnings above SGA, under current rules (control group) and in the return-to-work program (treatment group). The x-axis represents annual earnings in USD (NOK/USD \approx 6). The sample consists of individuals aged 18-49 (at the date of DI award), who were awarded DI benefits in Dec. 2003 & Jan. 2004. The vertical lines mark the maximum permitted annual earnings under current rules (E) and under the return-to-work program (e) (see Figure 1). The participation tax rate (PTR) is given by equation (5), and incorporates the DI system, income taxation, and dependent benefits. The earnings densities are calculated with a bin size of 0.5 SGA. The dots represent the unrestricted density in each bin, while the lines represent the estimated densities from local linear regressions.