DISCUSSION PAPER SERIES

IZA DP No. 6702

How Financial Incentives Induce Disability Insurance Recipients to Return to Work

Andreas Ravndal Kostøl
Magne Mogstad

July 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

How Financial Incentives Induce Disability Insurance Recipients to Return to Work

Andreas Ravndal Kostøl
Statistics Norway
and University of Bergen

Magne Mogstad
University College London,
Statistics Norway, IFS, ESOP and IZA

Discussion Paper No. 6702
July 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 6702
July 2012

ABSTRACT

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

Disability Insurance (DI) programs have long been criticized by economists for apparent work disincentives. Some countries have recently modified their programs 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. However, there is little empirical evidence of the effectiveness of programs that incentivize the return to work by DI recipients. 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 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.

JEL Classification: H53, H55, I18, J21

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

Corresponding author:

Magne Mogstad
Department of Economics
University College London
Gower Street
London WC1E 6BT
United Kingdom
E-mail: m.mogstad@ucl.ac.uk

* We thank Rolf Aaberge, Richard Blundell, Christian Brinch, Espen Bratberg, Paul Devereux, Tarjei Havnes, Torbjørn Hægeland, Per Johansson, Katrine Løken, Knut Røed, Kjetil Telle, Kjell Vaage, Till Von Wachter, and a number of seminar and conference participants for useful comments and suggestions.

1 Introduction

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 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 the 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 were 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 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

[1] These arguments are presented in several of the bills that have been brought to the floor of the U.S. Congress that have outlined 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 et al., 2009; Autor, 2011)

sharp discontinuity in DI policy therefore provides an attractive setting in which 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 implementaton, the return-to-work program has increased the labor force participation of DI recipients aged 18-49 by 8 percentage points. In comparison, less than 2 percent of the control group participated in the labor force, and labor force participation among rejected DI applicants of the same age was approximately 30 percent. Second, the return-to-work program produced a large increase in the earnings of DI recipients. These gains 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 males, highly educated individuals, and people in areas with low unemployment showing the strongest responses to the financial incentives.

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. This finding lends support to the external validity of our analysis of the consequences of providing financial incentives that encourage DI recipients to return to work.

Related literature: Our paper is primarily related to a fairly small body of 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 et al. (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. Our paper complements the existing evidence: the local randomized experiment that arises from the sharp discontinuity in DI policy provides transparent and credible evidence that financial incentives induce DI recipients to return to work. 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. However, we find no evidence of a decrease in the rate of exit from DI by current recipients as a result of the increase in program generosity.

Another related strand of the literature 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 encouraging return to work on the part of disability beneficiaries. Using U.S. data, Moore (2011) finds that removing drug and alcohol addictions from the group of disabling conditions substantially affected employment. Adam et al. (2011) evaluate the overall effects of the Pathways–to–Work program, which provided greater (financial and non-financial) support and created greater obligations to encourage work among new claimants seeking assistance through various illness- and disability-related schemes. The program is found to increase employment and accelerate the rate of outflow from these programs, but only for individuals who would have left the benefit rolls in less than a year in any case. Campolieti and Riddell (2012) study a policy change that was conducted in Canada, where the earnings limit that would trigger a continuing disability review was increased in the rest of the country to the figure used in the province of Quebec. The researchers find that the policy change increased the employment of DI recipients without inducing entry into the DI program.

Our paper is also related to a growing body of economics 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 the DI program and the Social Security system more generally. It also raises troubling questions regarding whether the program is being misused by recipients.[2] We show that many individuals who are deemed "totally and permanently" disabled by the Social Security Administration indeed have considerable

[2] Between 1985 and 2008, the share of non-elderly adults receiving DI benefits increased by 172% to 6% in the U.S. and by 126% to 7% in the UK. (Autor and Duggan, 2006; Moore, 2011; Burkhauser and Daly, 2012; OECD, 2010). The growth in the number of individuals who receive DI has important ramifications for economic growth and the public financing of social security systems. In 2007, DI payments constituted 1.7% of the GDP in the U.S. and 2.3% of the GDP across the European OECD-countries (OECD, 2010).

work capacity and that their labor supply is fairly elastic with respect to changes in financial work incentives.

Our findings complement the current evidence of 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 disability 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]

Outline: Section 2 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 on earnings, benefits, and income. Section 3 describes the data and discusses the empirical strategy. Section 4 presents the empirical results. Section 5 concludes.

2 Background

This section begins by providing an overview of the Norwegian DI program, describing who qualifies, how an individual applies and how the level of benefits is determined.[5] Next, we discuss the return-to-work program and its expected effects on DI recipients. We conclude this section with a comparison between the DI programs in the U.S. and in Norway.

[3]There is also a considerable body of evidence of how local labor market conditions and wages affect entry to DI (see e.g. Black et al. (2002), Bratberg (1999), Bratsberg et al. (2010), Rege et al. (2009)).

[4]See Chen and van der Klaauw (2008), French and Song (2009), Maestas et al. (2011), and Von Wachter et al. (2011).

[5]Throughout 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 vast 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.

2.1 The Norwegian DI program

In Norway, DI benefits are designed to provide partial earnings replacements 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 through employer- and employee-paid taxes.

The application process

To apply for DI benefits, an individual must submit an application to a Social Security Administration field office. The employees at the field office first review the non-medical criteria. Only adults below the (mandatory) retirement age of 67 years are eligible for DI. If the applicants 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. In situations in which the applicant has an impairment that is on the official listing of impairments, 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 field office concludes that the applicant cannot be expected to engage in any substantial gainful activity, a disability award is made.

In 2003, approximately 65 percent of applicants were awarded benefits at this first stage. The first stage in the application process takes at least three months and often lasts more than six months. If the original DI application is denied, an applicant may request a re-evaluation by the state Social Security Administration office. The applicant may also subsequently appeal through the courts if the re-evaluation is unsuccessful. Of those individuals who were initially denied benefits in 2003, approximately 75 percent requested re-evaluation. More than 44 percent of those who requested re-evaluation were awarded disability benefits within the next five years.

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 wages 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 the AIE increases so that low-wage workers replace a larger fraction of their earnings than do 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 claimant may die. Death accounted for 19 percent of all program exits in 2003. Third, the claimant 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.

2.2 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 (i.e., the control group) had to abide by the current rules. Below, we describe the two programs and use basic labor supply theory to make economic predictions.[7]

The programs

The treatment group and the control group are both faced with a 12-month waiting period after DI is awarded. If the DI recipient works during the waiting period, he or she becomes ineligible for DI benefits. After the waiting period, benefits are provided in full if the person's earnings are below the SGA threshold. Under the current rules and in the return-to-work program, the SGA threshold was set at \$1,050 of monthly earnings.[8] In both programs, a trial working period allows a person to have earnings above the SGA threshold for up to five consecutive years without risking reduced benefits should that person's earnings fall below the SGA threshold.

The key difference between the two programs is how benefits are reduced if earnings exceed the SGA threshold. Figure 1 illustrates this distinction. The figure presents stylized budget constraints in income-leisure space for the two programs. 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

[6]See Bratberg (1999) for a description of the formula determining the DI benefits.

[7]The description of the two programs is based on the Norwegian government's white paper from 2007 on disability insurance (NOU, 2007)

[8]Throughout this paper, all monetary figures are fixed at 2010 level. For the figures expressed in U.S. dollars (\$), we have used the following exchange rate: NOK/\$ = 6.

the period between July 2003 and February 2004. We set the wage equal to the average wage in a sample of workers who are not on DI but have observable characteristics similar to those of the DI recipients.[9] The line AG (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.

The budget constraint under the current rules is represented by $BCDEFG$. 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 constraints (represented by the line CD). At the budget segment DE, the 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, the earnings exceed the maximum permitted amount; this situation implies the absence of disability and therefore signals the individual's ineligibility for DI benefits.[10] This creates a second notch that is given by the line EF.

The budget constraint in the return-to-work program is $BC\overline{EF}G$. The program eliminates the first notch in the budget constraint under current rules and phases out benefits at budget segment $C\overline{E}$ at a rate of approximately \$0.6 for every \$1 in earnings. The program also extends 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 $\overline{EF}$.

The expected impacts

Table 1 summarizes the expected impacts of the transformation of the budget constraint from the current rules to the return-to-work program rules. We make the textbook labor supply model assumptions: The DI recipients choose hours of work at the given offered wage, and offered wages are constant. In particular, we ignore any human capital, search theoretical, or related issues. We also assume that preferences are convex and that leisure and income are normal goods.[11]

Table 1 consists of two panels. Panel (a) assumes that DI recipients can freely choose their working hours, in which case we only need to consider the set of points $\{B, C, H, G\}$ under the current rules to exhaust

[9] To compute the hourly wages, we use information from the 2005 Wage Statistics Survey 2005. On the basis of this sample of workers (who are not on DI), we regress hourly wages on the covariates included in Table 3 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.

[10] The permitted maximum amount equals 80 percent of annual earnings prior to disability.

[11] Note that labor supply theory makes predictions about hours worked. Assuming no change in the wages offered, this implies a prediction about earnings. Thus, the table includes a single prediction for hours/earnings, which is important because we observe earnings but not hours in our data.

all qualitatively possible responses to the return-to-work program. Panel (b) displays the expected impacts of the return-to-work program in a situation in which the DI recipients make discrete choices between no work, part-time work, and full time-work. Whether panel A or panel B applies depends on the extent of the institutional constraints on choices regarding working hours (e.g., in the form of fixed costs associated with working). Figure 1 illustrates this issue using the hours of work distribution in 2005 for a sample of workers who are not on DI. We can see that the vast majority of the employees were working either part-time (15 hours per week) or full-time (37.5 hours per week), indicating that there might be institutional constraints on the working hours chosen.

There is one key prediction that is common across the two panels: for those DI recipients who will work little, if at all, under current rules (i.e., locating at budget segment BC), labor supply theory predicts that the return-to-work program will (weakly) increase earnings and income and that it will (weakly) decrease benefits. 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 very few DI recipients who work under the current rules: in the control group, less than two percent of the DI recipients earn more than the SGA threshold. It is therefore likely that the return-to-work program will (weakly) increase average earnings and income and that it will (weakly) reduce average benefits.

Lastly, it should be noted that because the cut-off date for eligibility to the return-to-work program was set retroactively, any change in the number of individuals receiving DI benefits must come from decreases in exits (as at points H and G) rather than from non-recipients entering the program.

2.3 Comparison with the U.S. program

The DI programs in Norway and the U.S. are quite similar and feature only three important differences .[12]

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 graphing 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., the time trends are quite similar.[13] From 1961 to 2005, the rate of receipt increased from 2.2 to 10.3 percent in Norway and from 0.8 to 4.2

[12]Our 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).

[13]The cross-country difference in coverage by the DI program is unlikely to explain the entire discrepancy: although virtually all non-elderly adults are covered in Norway, more than 80 percent of all non-elderly adults are covered in the US. 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 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.[14] 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.

The second difference between the two programs is that the U.S. DI program is less generous. Figure 3 shows this difference by presenting stylized budget constraints in income-leisure space for the two DI programs. To compute the figures 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 compensation rate would be 31 percent in the U.S. program and 58 percent in the Norwegian program.[15] Although the two programs have similar SGA thresholds ($1,050 per month in Norway and $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 DI: This increases the tax rate associated with the first notch and eliminates the second notch, as compared to the Norwegian program.

Figure 3 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. Table 2 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

[14]See 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.

[15]For the sake of simplicity, our calculations for the compensation 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 compensation rate somewhat. In Norway, all citizens are eligible for health insurance through the Social Insurance System.

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 %).

3 Data and empirical strategy

This section begins by describing the data and sample selection. Next, we outline the RD design. Lastly, we examine the data, reporting summary statistics and providing a graphical representation of the empirical strategy.

3.1 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 information on educational attainment is based on annual reports from Norwegian educational establishments, whereas the other socio-economic data are collected from tax records and other administrative registers. The demographic information is from the Central Population Register, which is updated annually by the local population registries and verified by the Norwegian Tax Authority. The coverage and reliability of Norwegian registry data are considered to be exceptional, as illustrated by the fact that they received the highest rating in a data quality assessment conducted by Atkinson et al. (1995).

The outcomes that we consider are defined as follows. The first key outcome is labor force participation. As in Maestas et al. (2011), labor force participation is an indicator variable that is equal to one if annual earnings exceed the annual SGA threshold, set at approximately $12,500 per year. Unfortunately, we are unable to measure the labor supply at the intensive margin because we do not have data on working hours. [16] The second key outcome is annual gross earnings, which include wages and income from self-employment.

[16]Figure 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

We also consider the effect of the program on disposable income, incorporating earnings and benefits less 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 focuses on individuals who were awarded DI between July 2003 and February 2004. As discussed in more detail below, this sampling window is used to avoid confounding the effects of the program with consequences of another policy change that affected individuals who were awarded DI after March 2004. We also restrict the sample to individuals who were deemed totally and permanently disabled by the Social Security Administration because the return-to-work program was not targeted at partially or temporarily disabled individuals. Lastly, five (of nineteen) counties introduced a program in January 2004 that provided wage subsidies to firms that hired DI recipients. To identify the effects of the return-to-work program and to avoid confounding them with the effects of wage subsidies to firms, we exclude individuals who resided in these five counties in 2003 (reducing the sample by 25 percent).

Throughout the empirical analysis, we will partition the sample into two groups according to age. One reason is that the cost of working may be lower for younger DI recipients, who more often suffer from difficult-to-verify disorders such as mental illness and musculoskeletal disease (Autor and Duggan, 2006; Von Wachter et al., 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 et al., 2003; Von Wachter et al., 2011). In our main analysis, we will focus on individuals who were between the ages of 18 and 49 when they were awarded DI. In sub-section 4.4, we also report results for older DI recipients, those who were ages 50-61 when they were awarded DI.[17] Like French and Song (2009) and Von Wachter et al. (2011), we find that much is missed if we average the labor supply responses of younger and older DI recipients.

3.2 RD design

In January 2005, the Norwegian government introduced a program intended to induce the return to work among DI recipients. As described in detail above, the program allowed DI recipients in work to keep a significant fraction of their benefits. However, only recipients who were awarded DI before January 2004 were eligible for the program. Individuals awarded DI after that date were unable to enter the program by manipulating the award date. The local randomized experiment that arises from this sharp discontinuity in

return-to-work program on hours of work.

[17]The reason for the upper age restriction is that Norwegian workers can receive early retirement benefits beginning at age 62.

DI policy provides an attractive setting for analyzing the impact of the return-to-work program using a RD design that compares recipients awarded DI just before (i.e. the treatment group) and just after (i.e. the control group) January 2004.[18]

Treatment (T) for the return-to-work program is a deterministic function of the assignment variable, the date of the DI award (X), where the cut-off date c equals January 1 of 2004:

$$T \quad = \quad 1(X < c)$$

The identifying assumption is that individuals are unable to precisely control the assignment variable near the cut-off. This implies that variation in treatment near cut-off is random and

$$\tau^{RD} = lim_{\epsilon \uparrow 0} E\left[Y \mid X = c + \epsilon\right] - lim_{\epsilon \downarrow 0} E\left[Y \mid X = c + \epsilon\right]$$

equals the average effect of the return-to-work program among individuals who were awarded DI around the cut-off date.

The RD design can be implemented by using separate regressions on each side of c. For the sake of convenience, we transform X to $X - c$, so that the intercepts of the two regressions yield the value of the regression functions at c. The regression model on the left side of the cut-off date ($X < c$) is

$$Y = \alpha_l + f_l(X - c) + \varepsilon_l, \tag{1}$$

whereas the regression model on the right side of the cut-off date ($X > c$) is

$$Y = \alpha_r + f_r(X - c) + \varepsilon_r, \tag{2}$$

where f_r and f_l are unknown functional forms. The average effect of the return-to-work program can then be estimated as the difference between the regression intercepts on the two sides of the cut-off date

$$\hat{\tau}^{RD} = \hat{\alpha}_l - \hat{\alpha}_r. \tag{3}$$

To make inferences about the impact of the program, we pool the two regressions and follow Fuji et al.

[18]While the RD design was being used in applied economic research in studies such as Van Der Klaauw (2002) and Angrist and Lavy (1999), key issues of identification and estimation were formalized in the theoretical work of Hahn et al. (2001) and Lee (2008), and discussed in detail in Lee and Lemieux (2010).

(2009) in calculating the heteroskedastic robust standard errors.

To implement the RD design, we need to specify f_r and f_l and must 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 (2009). 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 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.

To increase our confidence in the RD design, we run a number of specification checks. We begin by including a set of (pre-determined) covariates, many of which are highly correlated with our post-assignment outcome variables of interest. As a consequence of a local randomized experiment, assignment to treatment near the cut-off date should be independent of pre-determined covariates. Thus, it is reassuring that the distribution of the covariates is stable around the cut-off date and that our results are robust to the inclusion of the covariates.

It is still possible that the RD design has been contaminated by (discontinuous) changes in the potential outcomes for DI recipients according to the month in which they were awarded DI. We assess the potential bias due to month of award effects with the following difference-in-differences version of the RD design

$$\tau^{DiD} = \tau^{RD} - \{lim_{\epsilon \uparrow 0} E\left[Y \mid X = d + \epsilon\right] - lim_{\epsilon \downarrow 0} E\left[Y \mid X = d + \epsilon\right]\},$$

where the cut-off date d equals January 1, 2003. The difference-in-differences specification exploits that recipients who were awarded DI just before and just after January 2003 were assigned to the return-to-work program. Significant differences in the post-assignment outcomes for the two groups should therefore be unrelated to the return-to-work program and should instead capture the effects of the month when the award was given. We implement the difference-in-differences version of the RD design by employing separate regressions on each side of the cut-off dates c and d: our first specification uses local linear regressions with a triangular kernel density and 2 months of bandwidth (this specification is denoted as RD-DiD), whereas our second specification uses a window of just one month and is therefore a standard difference-in-differences (DiD) model.

Lastly, we perform several robustness checks to the specification of f_r and f_l, and to the choice of bandwidth. That the estimates do not vary significantly from one specification to the other increases our confidence in the RD design.

3.3 Descriptive 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 provide a detailed description of how the outcomes, the covariates, and the density of the assignment variable vary around the cut-off date.

We begin by showing average labor force participation and mean earnings of recipients who were awarded DI in December of 2003 and January of 2004 (i.e., the treatment and control group in the FD model). As is evident from Table 3, there are substantial differences between the two groups in terms of labor market attachment after the commencement of the return-to-work program. The difference in labor force participation steadily increased from 2.2 percentage points in 2005 to 5 percentage points in 2007. In comparison, less than 2 percent of the control group participated in the labor force, and labor force participation among rejected applicants was approximately 30 percent. We also see that the average earnings from 2005 to 2007 are approximately twice as high in the treatment group as in the control group. In contrast to the stark difference in the post-assignment outcomes, the pre-determined covariates are fairly balanced across the two groups. We also see that there is virtually no difference between the numbers of DI recipients in the two groups. Overall, the pattern showed in Table 3 suggests substantial effects of the return-to-work program on the labor market attachment of DI recipients.

We next increase the windows on both sides of the cut-off date. Panel (a) in Figure 4 shows the mean labor force participation and average earnings for recipients who were awarded DI between July 2003 and June 2004; Panel (b) in Figure 4 shows the means for the same labor market outcomes for recipients who were awarded DI between July 2002 and June 2003; and Figure 5 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. Whereas the regression lines better illustrate the trends in the data and the size of the jumps at the cut-off dates, the unrestricted means indicate the underlying noise in the data. In each graph, the scale of the y-axis is set equal to $\pm.5$ standard deviation of the respective variable. By standardizing the y-axes in this way, we can easily compare the trends in the data and the sizes of the jumps at the cut-off dates across the graphs.

The graphical representation of the data mirrors the descriptive statistics reported in Table 3. Figure 4 shows evidence of a discontinuity in each outcome around January 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 2003. This finding is reassuring because there were no changes in assignment to the return-to-work program around this cut-off. Figure 5 shows that the covariates are generally quite stable around the cut-off date.

Note that the grey shaded areas in Figures 4 and 5 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 a DI application takes at least 3 months (and often up to or more than 6 months), recipients who were awarded DI before March 2004 were not affected by the introduction of temporary DI benefits. 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.[19]

As a consequence of the individuals being unable to manipulate the award date, the pre-determined covariates should have the same distribution just before and just after January 1 of 2004. Table 3 and Figure 5 support this conjecture. 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. McCrary (2008) proposes a simple two-step procedure for testing whether there is a discontinuity in the density of the assignment variable. In the first step, the assignment variable is partitioned into equally spaced bins, and frequencies are computed within those bins. In the second step, the frequency is used as a dependent variable in a local linear regression.

In Figure 6, we examine the density of the assignment variable and perform McCrary's two-step procedure. Panel (a) 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. Consistent with Table 3, the 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

[19] Our results are qualitatively the same and quantitatively similar if we extend the control group to include individuals 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.

January 2004, as manipulation of the application date would imply. As expected, the number of recipients awarded DI decreases in March 2004, as the introduction of temporary DI crowds out out some of the use of permanent DI benefits. Panel (b) displays the results from McCrary's two-step procedure. 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 -.00712 with a standard error of .151.

4 Empirical results

This section begins by showing how the return-to-work program affected the labor force participation and earnings of DI recipients. Next, we show that our findings are robust to a battery of specification tests. We then estimate the financial costs and benefits of the program, after which we examine the heterogeneity in the impact of the return-to-work program. We conclude this section with a comparison between the upper bound estimates for the counterfactual labor force participation rate among DI recipients in Norway and the U.S.

4.1 Labor force participation and earnings

FD RESULTS. Table 4 reports results from the FD specification. The first column directly compares the outcomes (and pre-determined covariates) for the recipients who were awarded DI in December 2003 (the treatment group) and January 2004 (the control group). The second column adds the set of pre-determined covariates to the FD model. The third column uses the DiD specification to control for the possible effects of the month when DI was awarded: the DiD model compares the difference between the outcomes (and pre-determined covariates) for the treatment and the control groups to the difference between the outcomes (and pre-determined covariates) for the recipients who were awarded DI in December 2002 and January 2003.

As expected, Table 4 shows that there are substantial differences between the labor market outcomes in the treatment and control 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, less than 2 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.

RD RESULTS. Table 5 reports the results obtained from the RD specification, 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 Table 5, 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: for 2005, the estimate is approximately 3 percentage points, and for 2007, it exceeds 8 percentage points. We can also see that the RD model estimates the increase in earnings to be approximately 110 percent.

COVARIATES AND MONTH EFFECTS. In comparing the estimates across the first and the second columns of each table, one can see that the RD and the FD results are quite similar when the covariates are included. Throughout the paper, we allow for a flexible functional form by including a large set of dummies for different values of the covariates. By comparing the estimates in the first and the third column of each table, we can see that month of the award effects do not confound the conclusions drawn about whether financial incentives encourage DI recipients to return to work.

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 an useful way to gain precision, especially in the subsample analysis where the sample size is relatively small (see e.g. Froelich, 2007).[20]

AGGREGATE LABOR FORCE PARTICIPATION ELASTICITY. Tables 4 and 5 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 labor force participation elasticities with respect to the participation tax rates (PTR). In situations

[20]The results are qualitatively the same and quantitatively similar if we drop covariates or use the difference-in-differences specification. The results are available upon request.

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 et al., 2006; Keane, 2011).

The PTR summarizes how the tax and benefit systems affect the financial gain from work and is defined as

$$PTR = 1 - \frac{I - I_0}{W}$$

where I_0 denotes the amount of disposable income if one is not working, whereas I and W denote the disposable income and earnings if one is working. If the PTR is greater than 1, the DI recipient would be worse off while working then while not working; a PTR equal to 1 means that there is no financial reward for work; and a PTR of zero means that the financial reward for work is equal to one's gross earnings.

We then follow Gruber (2000) and French and Song (2009) in using the arc[21] elasticity of labor force participation, defined as

$$\epsilon = \frac{\triangle LFP/\overline{LFP}}{\triangle PTR/\overline{PTR}}, \tag{4}$$

where $\triangle LFP$ denotes the difference between the labor force participation of the treatment group and that of the control group, $\overline{LFP}$ is the mean employment rate in the treatment and the control group, $\triangle PTR$ denotes the difference between the treatment and the control group in terms of PTR, and $\overline{PTR}$ is the mean PTR in the treatment and the control group.

To compute the aggregate labor force participation elasticity ϵ, we need to decide how to measure the PTR. We begin by setting 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.[22] This allows us to compute a measure of PTR for any given earnings level (k), given by

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

Figure 7 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.

[21]The arc elasticity uses the average of the treatment and control group as the base, and is therefore symmetric with respect to the labor force participation and PTR levels in the two groups. Thus, the arc elasticity is less sensitive to changes in the very low labor force participation rate of the control group, as compared to an elasticity using the control group mean as the base.

[22]The aggregate labor force participation elasticities barely move if we instead set I_0 equal to zero.

We then define

$$\triangle PTR = \sum_{k>SGA} [E(PTR_k \mid Treatment) - E(PTR_k \mid control)]p_k \tag{6}$$

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

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

where dk is set equal to half of the SGA amount ($6,250 per year). This means that $\triangle PTR$ 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.[23] Figure 7 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 $\triangle PTR$. As expected, the return-to-work program also had the largest impact on earnings levels near the SGA.

Table 6 shows the labor force participation elasticities implied by the estimated effects on labor force participation. The results suggest that the labor force participation of DI recipients increased by between 1.9 and 2.5 percent for each one-percent decrease in the PTR. The relatively large aggregate elasticities are broadly consistent with previous studies of labor supply and income taxation in situations with notches (or kinks) in the budget set (see e.g. Eissa et al., 2006; Kleven and Kreiner, 2006).

4.2 Robustness analysis

Table 7 reports the results of a placebo test that moves the cut-off date from January 2004 to January 2003. The placebo test exploits that recipients awarded DI just before and just after January 2003 were assigned to the return-to-work program. Significant differences in the post-assignment outcomes of the two groups is therefore a sign of misspecifation of the potential outcome curves, rather than a true program impact. It is therefore reassuring that neither the FD model nor the RD model suggests significant differences between the outcomes of the two groups.

Table 8 reports the results of several specification checks of the RD model. That the estimates do not

[23] The aggregate labor force participation elasticities change little if we instead weight by the (conditional) density of the earnings by those individuals in the treatment group who work.

vary significantly from one specification to the other increases our confidence in the RD design. We begin by extending 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. We next 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.

Lastly, we constrain $f(X-c)$ 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 make $f(X-c)$ a common polynomial function of X.

4.3 Financial costs and benefits

Advocates of programs such as the "$1 for $2" offset program 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 9 shows the results.

Our estimates indicate that the return-to-work program was successful in increasing disposable income because the gains in earnings exceeded the loss in benefits and the additional taxes paid. It is also evident that the increase in disposable income was accompanied by a decrease in program costs that ranged from 3.5 to 5 percent of the costs associated with DI under the current rules. The reduction in costs is attributable to a significant decrease in benefits and a smaller increase in taxes paid. Because we find no effect of the return-to-work program on the rate of exit from DI,[24] it appears that the decrease in costs occurs because of an increase in part-time work by DI recipients rather than a reduction in DI rolls.

4.4 Heterogeneity

Tables 4 and 5 demonstrated that many DI recipients have a considerable capacity to work and can be effectively encouraged to work if financial incentives are provided. However, these findings should be interpreted

[24]The estimated effects on the exit rate are available upon request. Both in the return-to-work program and under current rules, very few individuals exit the DI program because they earn more than the permitted maximum amount; the vast majority of exits from DI occur because the individual in question reaches the retirement age.

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.

As is evident from the second column in the top panel in Table 10, 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 explanation for this is that the cost of working may be lower for younger DI recipients, who tend to suffer more often from difficult-to-verify disorders such as mental illness and musculoskeletal disease (Autor and Duggan, 2006; Von Wachter et al., 2011). Another explanation is that the gains from returning to the labor force may be greater for younger DI recipients because they have more years to potentially work prior to their retirement (Bound et al., 2003; Von Wachter et al., 2011).

Table 10 also shows that within the group of younger DI recipients, the estimates vary substantially, with the largest effects of the program on males, high educated, and people in low unemployment areas. Figure 8 summarizes this heterogeneity: For each subgroup, it plots the average $\triangle LFP$ and the average ϵ over the period 2005-2007. There are clear patterns in the two graphs: highly educated people, males, and people in areas with low unemployment tend to be relatively elastic with respect to changes in financial incentives. Figure 8 also displays the regression lines between the subgroup means of $\triangle LFP$ and ϵ. The strong association suggests that the heterogeneity in the effects of the program on labor force participation is driven by differences between individuals' levels of responsiveness to financial incentives (rather than by differences in $\triangle PTR$ or the bases, $\overline{LFP}$ and $\overline{PTR}$).

4.5 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. The comparison puts the magnitude of the RD estimates into perspective, and sheds light on the likely 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. Bound's empirical analysis has recently been replicated and extended. Chen and van der Klaauw (2008) apply his approach to data on DI applicants from the 1990s. Their estimates indicate that the labor force participation rate among DI recipients in the 1990s would have been at most 20 percentage points higher had they not received benefits. Using data from 1978 to 2004, Von Wachter et al. (2011) replicate Bound's analysis for male applicants aged 45-64 and show that his main conclusion is quite stable over time and is robust to several alternative specifications. They next extend Bound's analysis to male applicants ages 30-44 and estimate that the labor force participation rate of DI recipients is at most 32.9 percent. Lastly, French and Song (2009) and Maestas et al. (2011) also present estimates based on the Bound approach. Their findings suggest that the counterfactual labor force participation rate among DI recipients would be between 24 and 27 percent.[25]

Although there is substantial 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 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. Table 11 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 November or December of 2003, individuals who were awarded DI in November or December of 2003 (the treatment group), and individuals who were awarded DI in January and February of 2004 (the 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. Following previous studies that used the Bound approach, we exclude rejected applicants who successfully reapply at a later date.

We see that the labor force participation rate is 22-25 (25-30) percentage points higher among rejected applicants than in the treatment group (the control group). In the Bound approach, these differences are

[25] Chen and van der Klaauw (2008), French and Song (2009) and Maestas et al. (2011) also use quasi-experimental approaches to address the concern that those who are denied benefits are potentially different from those who are allowed benefits. The findings are mixed and suggest that the disincentive effect of DI benefits is quite heterogeneous.

interpreted as upper bound estimates of the counterfactual labor force participation rate among DI recipients. However, there are some significant differences between 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 5-8 percentage points. These estimates are quite similar to the recent US evidence, particularly the results for younger DI recipients reported in Von Wachter et al. (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 lend support to the external validity of the RD estimates.

5 Conclusion

Autor and Duggan (2006) argue that the liberalization of the screening process and the movement away from physically exertive work have blurred any sharp divide that may once have existed between those who are 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. Thus, any increase in the number of individuals receiving DI benefits must be a result of decreases in exits rather than of increases in new recipients. We do, however, find no effect of the increase in program generosity on the exit rate from DI by current recipients.

Lastly, we advise readers to exercise 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 lend some support to the external validity of our empirical analysis.

References

Adam, S., Bozio, A., and Emmerson, C. (2011). Reforming disability insurance in the uk: Evaluation of the patways to work programme. *IFS Working Paper*.

Angrist, J. and Lavy, V. (1999). Using maimondies rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114:533–575.

Atkinson, A., Rainwater, L., and Smeeding, T. (1995). Income distributions in oecd countries: evidence from the luxembourg income study. *OECD Publications and Information Center*.

Autor, D. (2011). The unsustainable rise of the disability rolls in the united states: Causes, consequences, and policy options. *NBER*, No. 17697.

Autor, D. H. and Duggan, M. G. (2003). The rise in the disability rolls and the decline in unemployment. *Quarterly Journal of Economics*, 118(1):157–205.

Autor, D. H. and Duggan, M. G. (2006). The growth in the social security disability rolls: A fiscal crisis unfolding. *Journal of Economic Perspectives*, 20(3):71–96.

Benitez-Silva, H., Buchinsky, M., and Rust (2010). Induced entry effects of a $1 for $2 offset in SSDI benefits. *Working Paper SUNY-Stony Brook*.

Black, D., Daniel, K., and Sanders, S. (2002). The impact of economic conditions on participation in disability programs: Evidence from the coal boom and bust. *The American Economic Review*, 92(1):27–50.

Bound, J. (1989). The health and earnings of rejected disability insurance applicants. *American Economic Review*, 79(3):482–503.

Bound, J., Burkhauser, R., and Nichols, A. (2003). Tracking the household income of ssdi and ssi applicants. In Polachek, S. W., editor, *Research in Labor Economics*, volume 22, pages 113–158. Emerald Group Publishing.

Bratberg, E. (1999). Disability retirement in a welfare state. *Scandinavian Journal of Economics*, 101(1):97–114.

Bratsberg, B., Fevang, E., and Roed, K. (2010). Disability in the welfare state: An unemployment problem in disguise? *IZA Discussion Paper*, No 4897.

Burkhauser, R. V. and Daly, M. C. (2012). Social security disability insurance: Time for fundamental change. *Journal of Policy Analysis and Management*, Forthcoming.

Campolieti, M. and Riddell, C. (2012). Disability policy and the labor market: Evidence from a natural experiment in canada, 1998-2006. *Journal of Public Economics*, 96:306–316.

Chen, S. and van der Klaauw, W. (2008). The work disincentive effects of the disability insurance program in the 1990s. *Journal of Econometrics*, 142(2):757–784.

Eissa, N., Kleven, H. J., and Kreiner, C. T. (2006). Welfate effects of tax reform and labor supply at the intensive and extensive margins. In Agell, J. and Sørensen, P. B., editors, *Tax policy and labor market performance*. MIT Press.

French, E. and Song, J. (2009). The effect of disability insurance receipt on labor supply. *FRB of Chicago Working Paper No. 2009-05*.

Froelich, M. (2007). Regression discontinuity design with covariates. *University of St. Gallen Department of Economics Working Paper Series 2007*, WP 2007-32.

Fuji, D., Imbens, G., and Kalyanaraman, K. (2009). Notes for matlab and stata regression discontinuity software. *Harvard University Note*.

Gruber, J. (2000). Disability insurance benefits and labor supply. *Journal of Political Economy*, 108(6):1162–1183.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression discontinuity design. *Econometrica*, 69(1):201–209.

Hoynes, H. W. and Moffitt, R. (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.

Imbens, G. W. and Kalyanaraman (2009). Optimal bandwidth choice for the regression discontinuity estimator. *NBER Working paper*, No. 14726.

Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature*, 49(4):961–1075.

Kleven, H. J. and Kreiner, C. T. (2006). The marginal cost of public funds: Hours of work versus labor force participation. *Journal of Public Economics*, 90(10-11):1955 – 1973.

Lee, D. S. (2008). Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics*, 142:675–697.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

Livermore, G., Roche, A., and Prenovitz, S. (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, N., Mullen, K., and Strand, A. (2011). Does disabilityt insurance discourage work? *RAND Working paper*.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142:698–714.

Milligan, K. S. and Wise, D. A. (2011). Social security and retirement around the world: Mortality, health, employment and disability insurance participation and reforms. *Harvard Kennedy School Working paper*.

Moore, T. J. (2011). The employment effects of terminating disability benefits: Insights from removing drug addictions as disabling conditions. *Working Paper University of Maryland*.

NOU (2007). Ny uførestønad og ny alderspensjon til uføre. White Paper NOU 2007:4 (In Norwegian).

OECD (2010). *Sickness, Disability and Work: Breaking the Barriers. A synthesis of findings across OECD countries*. OECD Publishing.

Rege, M., Telle, K., and Votruba, M. (2009). The effect of plant downsizing on disability pension utilization. *Journal of the European Economic Association*, 7(4):754–785.

Stapleton, D., Livermore, G., Thornton, C., O'Day, B., Weathers, R., Harrison, K., O'Neil, S., Martin, E. S., Wittenburg, D., and Wright, D. (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.

Van Der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression discontinuity approach. *International Economic Review*, 43(4):1249–1287.

Von Wachter, T., Song, J., and Manchester, J. (2011). Trends in employment and earnings of allowed and rejected applicants to the social security disability insurance program. *American Economic Review*, Forthcoming.

Table 1: Expected effects of return-to-work Program

Location if assigned to current rules	Location on return-to-work program	Effect of return-to-work program on: Hours/Earnings	Benefits	Income
		(a) Free choice of hours		
B	B	0	0	0
C	C	0	0	0
C	$C\overline{E}$	+	-	+
H	$C\overline{E}$	-	+	+
G	G	0	0	0
G	$C\overline{E}$	-	+	-
		(b) Fixed set of hours		
B	B	0	0	0
B	Part or full time	+	-	+
Part time	Part time	0	+	+
Part-time	Full time	+	-	+
Full time	Full time	0	+	+
Full-time	Part time	-	+	+

Note: This table summarizes the expected impacts of the return-to-work program on earnings, benefits and income, based on Figure 1. Panel (a) assumes that DI recipients can freely choose hours of work. Panel (b) assumes that the DI recipients make discrete choices between either no work, part-time work (15 hours per week), or full time work (37.5 hours per week).

Table 2: Characteristics of DI recipients in Norway and the U.S.

Characteristics	Norway	US
Age at award decision	51.69	47.38
Avg. earnings, 3-5 years before ($)	27 143	25 503
Musculoskeletal diseases (%)	32.8	35.1
Mental disorders (%)	28.6	22.3
Cardiovascular system (%)	9.8	8.9
Other diagnosis (%)	28.8	33.7
Obs.	18 382	1 535 610

Note: The Norwegian sample consists of individuals aged 18-66 (at the date of DI award), who were awarded DI benefits in 2003. The U.S. statistics come from Maestas et al. (2011), and cover individuals aged 18-64 (at the date of DI award), who applied for and were awarded DI benefits in the period 2005-2006. Maestas et al. (2011) define average earnings in the U.S. as mean earnings over the last 3-5 years before DI award. We define average earnings in Norway as mean earnings over the last 4 years before DI award.

Table 3: Summary statistics

	Treatment (Dec. 2003)	Control (Jan. 2004)	Rejected Applicants (Nov. & Dec. 2003)
		Outcomes	
LFP(2005)	0.036	0.014	0.262
	(0.188)	(0.118)	(0.441)
LFP(2006)	0.050	0.019	0.307
	(0.218)	(0.135)	(0.462)
LFP(2007)	0.073	0.019	0.324
	(0.260)	(0.135)	(0.469)
Avg. earnings	2367	1241	13856
(2005-2007)	(6092)	(4344)	(22241)
		Characteristics	
Age at DI award/rejection	38.2	38.5	38.8
	(9.3)	(9.7)	(7.3)
Male	0.45	0.50	0.47
	(0.50)	(0.50)	(0.50)
Years of schooling	10.8	10.4	10.1
	(2.2)	(3.5)	(3.5)
Experience	13.9	14.1	10.1
	(9.0)	(10.1)	(8.2)
AIE	38279	38568	32949
	(14727)	(17799)	(14051)
Local unemployment rate	0.023	0.024	0.024
	(0.006)	(.008)	(.007)
Local DI rate	0.103	0.098	0.096
	(0.026)	(0.025)	(0.023)
Number of children	0.94	0.97	1.01
	(1.09)	(1.15)	(1.26)
Musculosceletal diseases	0.223	0.209	0.4
	(.417)	(.408)	(.491)
Mental disorders	0.445	0.405	0.298
	(.498)	(.492)	(.458)
Obs	220	215	225

Note: The first and second column show mean and standard deviation (in parentheses) of the outcomes and characteristics for individuals aged 18-49 (at the date of DI award), who were awarded DI in Dec. 2003 and Jan. 2004. The third column shows mean and standard deviation (in parentheses) of the outcomes and characteristics for individuals aged 18-49 (at the date of DI rejection), whose DI application was rejected in Nov. and Dec. 2003. Rejected applicants who successfully reapply during the following five years are excluded from the sample. Labor force participation (LFP) is equal to one if annual earnings exceed the SGA threshold. Average earnings is the mean earnings over the period 2005-2007. All individual characteristics are measured prior to disability award. Experience is the number of years with earnings above SGA. Average indexed earnings (AIE) summarize the earnings history of the DI recipients (the formula is given in footnote 2.1). 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 adult population (aged 18-66) receiving DI benefits in the municipality of residence 2-3 years prior to DI benefit award. 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: musculosceletal diseases and mental disorders.

Table 4: Main Results, FD Model

Dependent variable:	FD	FD w/c	DiD	Comparison means	
				January 2004	**Rejected applicants**
LFP(2005)	0.022	0.033**	0.020	0.014	0.262
	(0.015)	(0.016)	(0.021)	(0.118)	(0.441)
LFP(2006)	0.031*	0.033*	0.029	0.019	0.307
	(0.017)	(0.018)	(0.022)	(0.135)	(0.462)
LFP(2007)	0.054***	0.053**	0.056**	0.019	0.324
	(0.02)	(0.022)	(0.026)	(0.135)	(0.469)
Avg. earnings	1126**	1247**	1195*	1241	13856
(2005-2007)	(508)	(535)	(669)	(4344)	(22241)

Dependent variable:	FD	DiD	Comparison means	
			January 2004	**Rejected applicants**
Age at DI award	-0.30	0.89	38.5	38.8
	(0.91)	(1.27)	(9.7)	(7.3)
Male	-0.052	-0.003	0.50	0.47
	(.048)	(0.06)	(0.50)	(0.50)
Years of Schooling	0.32	0.41	10.4	10.1
	(0.28)	(0.38)	(3.5)	(3.5)
Experience	-0.21	2.0	14.1	10.1
	(0.92)	(1.30)	(10.1)	(8.2)
AIE	-290	1726	38568	32949
	(1565)	(2178)	(17799)	(14051)
Local unemployment rate	-0.001	0.000	0.024	0.024
	(0.001)	(0.001)	(0.008)	(0.007)
Local DI rate	0.004	0.006*	0.098	0.096
	(0.003)	(0.003)	(0.025)	(0.023)
Number of Children	-0.031	-0.028	0.967	1.013
	(0.11)	(0.15)	(1.14)	(1.26)
Musculosceletal system	0.013	-.015	0.209	0.400
	(0.040)	(0.054)	(0.408)	(0.491)
Mental disorders	0.041	-.03	0.405	0.298
	(0.048)	(0.065)	(0.492)	(0.458)
Obs	435	937	215	225

*** significant at 1% level, ** significant at 5% level, *significant at 10% level.

Note: This table displays estimates from the FD model, without any controls (FD), with covariates (FD w/c), and with controls for month effects (DiD). Standard errors (in parentheses) are robust to heteroscedasticity. Labor force participation (LFP) is equal to one if annual earnings exceed the SGA threshold. Average earnings is the mean earnings over the period 2005-2007. All covariates are measured prior to disability award, and described in Table (3). We allow for a flexible functional form by including a large set of dummies for different values of the covariates. FD and FD w/c use the sample of individuals aged 18-49 (at the date of DI award), who were awarded DI in Dec. 2003 and Jan. 2004. DiD uses the sample of individuals aged 18-49 (at the date of DI award) who were awarded DI in Dec 2002, Jan. 2003, Dec. 2003 or Jan. 2004. The second to last column shows mean and standard deviation (in parentheses) of the outcomes and characteristics for individuals aged 18-49 (at the date of DI award), who were awarded DI in Jan. 2004. The last column shows mean and standard deviation (in parentheses) of the outcomes and characteristics for individuals aged 18-49 (at the date of DI rejection), whose DI application was rejected in Nov. and Dec. 2003. Rejected applicants who successfully reapply during the following five years are excluded from the sample.

Table 5: Main Results, Regression discontinuity Model

Dependent variable:	**RD**	**RD w/c**	**RD-DiD**	**Comparison means**	
				Jan. & Feb. 2004	**Rejected applicants**
LFP(2005)	0.028	0.038	0.027	0.018	0.262
	(0.024)	(0.025)	(0.033)	(0.134)	(0.441)
LFP(2006)	0.039	0.042*	0.042	0.02	0.307
	(0.027)	(0.026)	(0.036)	(0.142)	(0.462)
LFP(2007)	0.087***	0.085***	0.082**	0.034	0.324
	(0.031)	(0.031)	(0.041)	(0.182)	(0.469)
Avg. earnings	1630**	1644**	1832*	1551	13856
(2005-2007)	(799)	(781)	(1080)	(5033)	(22241)

Dependent variable:	**RD**	**RD-DiD**	**Comparison means**	
			Jan. & Feb. 2004	**Rejected applicants**
Age at DI award	-0.20	1.52	38.4	38.8
	(1.44)	(2.00)	(9.7)	(7.3)
Male	-0.08	-0.01	0.50	0.47
	(0.08)	(0.10)	(0.50)	(0.50)
Years of Schooling	0.29	0.22	10.5	10.1
	(0.44)	(0.60)	(3.1)	(3.5)
Experience	-0.43	2.79	13.9	10.1
	(1.45)	(2.00)	(10.0)	(8.2)
AIE	-1123	1810	38013	32949
	(2486)	(3435)	(17686)	(14051)
Local unemployment rate	-0.002	-0.0004	0.024	0.024
	-0.001	(0.0014)	(0.008)	(0.007)
Local DI rate	0.005	0.0079	0.098	0.096
	-0.004	(0.005)	(.024)	(.023)
Number of Children	-0.08	-0.020	0.90	1.01
	(0.17)	(0.23)	(1.13)	(1.26)
Musculosceletal system	0.026	-0.001	0.22	0.4
	(0.062)	(0.084)	(0.415)	(0.491)
Mental disorders	0.041	-0.116	0.386	0.298
	(0.075)	(0.102)	(0.487)	(0.458)
Obs	897	1930	440	225

*** significant at 1% level, ** significant at 5% level, *significant at 10% level.

Note: This table displays estimates from the RD model, without any controls (RD), with covariates (RD w/c), and with controls for month effects (RD-DiD). We use a local linear regression with a triangular kernel density and 2 months of bandwidth on each side of the cut-off date. Standard errors (in parentheses) are robust to heteroscedasticity. Labor force participation (LFP) is equal to one if annual earnings exceed the SGA threshold. Average earnings is the mean earnings over the period 2005-2007. All covariates are measured prior to disability award, and described in Table (3). We allow for a flexible functional form by including a large set of dummies for different values of the covariates. RD and RD w/c use the sample of individuals aged 18-49 (at the date of DI award), who were awarded DI between Nov. 2003 and Feb. 2004. RD-DiD uses the sample of individuals aged 18-49 (at the date of DI award) who were awarded DI between Nov. 2003 and Feb. 2004 or between Nov. 2002 and Feb. 2003. The second to last column shows mean and standard deviation (in parentheses) of the outcomes and characteristics for individuals aged 18-49 (at the date of DI award), who were awarded DI in Jan. 2004. The last column shows mean and standard deviation (in parentheses) of the outcomes and characteristics for individuals aged 18-49 (at the date of DI rejection), whose DI application was rejected in Nov. and Dec. 2003. Rejected applicants who successfully reapply during the following five years are excluded from the sample.

Table 6: Aggregate Labor Force Participation Elasticities

	FD model			RD model		
Year	**2005**	**2006**	**2007**	**2005**	**2006**	**2007**
$\triangle LFP$	0.033	0.033	0.053	0.038	0.042	0.085
LFP in control group	0.014	0.019	0.019	0.018	0.020	0.034
$\overline{LFP}$	0.031	0.036	0.046	0.037	0.041	0.077
$\triangle PTR$	-0.34	-0.40	-0.46	-0.34	-0.40	-0.35
PTR in control group	1.01	1.02	1.02	1.01	1.01	1.02
$\overline{PTR}$	0.84	0.81	0.79	0.83	0.80	0.84
Easticity(ϵ)	-2.2	-1.9	-2.3	-2.3	-2.3	-2.5

Note: This table displays aggregate labor force participation elasticities for the sample of individuals aged 18-49 at the date of DI award. In the FD model (described in Table 4), the sample is restricted to individuals awarded DI benefits during the period December 2003 - January 2004. In the RD model (described in Table 5), the sample is resticted to individuals awarded DI benefits during the period November 2003 - February 2004. The differences in labor force participation (ΔLFP) are based on the estimates from the FD model and the RD model with controls (see the second column of Tables 4-5). The participation tax rate (PTR) is given by equation (5), and incorporates the DI system, income taxation, and dependent benefits. The weighted sum of the differences in PTR between the treatment and the control group (ΔPTR) is given by equation (6). The elasticity (ϵ) is given by equation (4). The bases $\overline{PTR}$ and $\overline{LFP}$ are the weighted averages of PTR and LFP in the treatment and the control group.

Table 7: Placebo test

Dependent variable:	**FD**	**DiD**	**RD**	**RD-DiD**
LFP(2005)	0.003	0.0052	0.003	0.011
	(0.014)	(0.0221)	(0.022)	(0.035)
LFP(2006)	0.003	0.0103	0.009	0.012
	(0.013)	(0.0257)	(0.022)	(0.042)
LFP(2007)	-0.002	-0.007	0.006	0.006
	(0.017)	(0.028)	(0.026)	(0.044)
Avg. Earnings	-68	-166.6	-202	-354
(2005-2007)	(440)	(770)	(727)	(1268)
Obs	502	970	1033	1968

*** significant at 1% level, ** significant at 5% level, *significant at 10% level.

Note: This table displays estimates of the FD and the RD model when we move the cut-off date from January 2004 to January 2003. Standard errors (in parentheses) are robust to heteroscedasticity. Labor force participation (LFP) is equal to one if annual earnings exceed the SGA threshold. Average earnings is the mean earnings over the period 2005-2007. The FD model uses the sample of individuals aged 18-49 (at the date of DI award), who were awarded DI in Dec. 2002 and Jan. 2003. DiD uses the sample of individuals aged 18-49 (at the date of DI award) who were awarded DI in Dec 2001, Jan. 2002, Dec. 2002 and Jan. 2003. The RD model uses the sample of individuals aged 18-49 (at the date of DI award), who were awarded DI between Nov. 2002 and Feb. 2003. RD-DiD uses the sample of individuals aged 18-49 (at the date of DI award) who were awarded DI between Nov. 2002 and Feb. 2003 and between Nov. 2001 and Feb. 2002.

Table 8: ROBUSTNESS ANALYSIS

PANEL (A)	Local linear regressions		Separate linear parametric trends			
	Oct. 2003-Feb. 2004		Nov. 2003-Feb. 2004		Oct. 2003-Feb. 2004	
LFP(2005)	0.031	0.035	0.028	0.043*	0.033	0.040*
	(0.021)	(0.021)	(0.024)	(0.025)	(0.021)	(0.022)
LFP(2006)	0.039	0.037*	0.039	0.042	0.038	0.042*
	(0.024)	(0.022)	(0.027)	(0.027)	(0.024)	(0.024)
LFP(2007)	0.079***	0.071***	0.087***	0.079**	0.074***	0.064**
	(0.027)	(0.026)	(0.031)	(0.032)	(0.027)	(0.028)
Avg. earnings	1551**	1372**	1630**	1728**	1494**	1466*
(2005-2007)	(712)	(676)	(814)	(828)	(745)	(755)
Controls	NO	YES	NO	YES	NO	YES
Obs	1106	1106	897	897	1106	1106

PANEL (B)	Common parametric trends (Oct. 2003 - Feb. 2004)					
	Linear		Quadratic		Cubic	
LFP(2005)	0.033*	0.039**	0.032	0.041*	0.023	0.041
	(0.019)	(0.020)	(0.022)	(0.023)	(0.029)	(0.030)
LFP(2006)	0.042*	0.040*	0.037	0.043*	0.042	0.04
	(0.022)	(0.022)	(0.025)	(0.025)	(0.033)	(0.033)
LFP(2007)	0.069***	0.058**	0.079***	0.069**	0.097***	0.085**
	(0.025)	(0.025)	(0.029)	(0.029)	(0.038)	(0.039)
Avg. earnings	1310*	1218*	1575**	1586**	1704*	1814*
(2005-2007)	(683)	(691)	(783)	(795)	(1033)	(1049)
Controls	NO	YES	NO	YES	NO	YES
Obs	1106	1106	1106	1106	1106	1106

*** significant at 1% level, ** significant at 5% level, *significant at 10% level.

Note: This table displays results from robustness checks of the RD model. Labor force participation (LFP) is equal to one if annual earnings exceed the SGA threshold. Average earnings is the mean earnings over the period 2005-2007. Standard errors (in parentheses) are robust to heteroscedasticity. All covariates are measured prior to disability award, and described in Table 3. When including covariates, we allow for a flexible functional form by including a large set of dummies for different values of the covariates. The sample is restricted to individuals aged 18-49 at the date of DI award, during the period described in each column. Panel A allows for seperate trends on each side of the cut-off date: The first two columns use local linear regressions with a kernel density; the four next columns use linear trends in the assignment variable. Panel B restricts the trends to be the same on both sides of the cut-off date: The first two columns use a linear trend in the assigment variable; the two next columns use a quadratic trend in the assignment variable; the last two columns use a cubic trend in the assignment variable.

Table 9: Financial costs and benefits

	Estimates		Comparison means	
Dependent variable:	FD-model	RD-model	Jan. & Feb 2004	Rejected Applicants
Avg. Earnings	1247**	1644**	1551	13856
(2005-2007)	(535)	(782)	(5032.6)	(22241)
Avg. Benefits	-798**	-1155**	33770	N/A
(2005-2007)	(359)	(536)	(8080)	
Avg. Tax	201	263	5234	4972
(2005-2007)	(156)	(229)	(2790)	(6204)
Avg. Disposable Income	248	226	30087	24101
(2005-2007)	(279)	(422)	(6217)	(19283)
Avg. Program Cost	-1000**	-1418**	28535	N/A
	(396)	(580)	(6169)	
Obs	435	897	440	225

*** significant at 1% level, ** significant at 5% level, *significant at 10% level.

Note: This table displays the impact of the return-to-work program on earnings, benefits, taxes, disposable income and program costs over the period 2005-2007. The sample consists of individuals aged 18-49 at the date of DI award. The estimates are based on the FD model and the RD model with controls. Standard errors (in parentheses) are robust to heteroscedasticity. In the FD model (described in Table 4), the sample is restricted to individuals awarded DI benefits during the period December 2003 - January 2004. In the RD model (described in Table 5), the sample is restricted to individuals awarded DI benefits during the period November 2003 - February 2004. Taxes include tax on earnings and benefits. Program cost equals benefits minus taxes. Disposable income equals earnings plus benefits minus taxes. The second to last column shows mean and standard deviation (in parentheses) of the outcomes for individuals aged 18-49 (at the date of DI award), who were awarded DI in Jan. and Feb. 2004. The last column shows mean and standard deviation (in parentheses) of the outcomes for individuals aged 18-49 (at the date of DI rejection), whose DI application was rejected in Nov. and Dec. 2003. Rejected applicants who successfully reapply during the following five years are excluded from the sample.

Table 10: Subsample Analysis

Panel (a)	Age groups				Gender			
Dependent variable:	Baseline		Aged 50-61		Males		Females	
	FD model	RD model	FD model	RD model	FD model	RD model	FD model	RD model
LFP(2005)	0.031*	0.038	-0.0004	-0.007	0.055*	0.073*	0.0003	-0.011
	(0.016)	(0.025)	(0.011)	(0.016)	(0.031)	(0.039)	(0.013)	(0.026)
LFP(2006)	0.032*	0.0431*	-0.003	-0.004	0.054*	0.072*	0.014	0.025
	(0.018)	(0.026)	(0.010)	(0.015)	(0.032)	(0.042)	(0.021)	(0.028)
LFP(2007)	0.055**	0.085***	-0.016	-0.031*	0.074**	0.095**	0.038	0.063*
	(0.021)	(0.031)	(0.011)	(0.017)	(0.033)	(0.044)	(0.030)	(0.038)
Obs	435	897	972	2132	205	439	230	458

Panel (b)	Education groups				Local labor market conditions			
Dependent variable:	Low Education		High Education		Low Unemployment		High Unemployment	
	FD model	RD model	FD model	RD model	FD model	RD model	FD model	RD model
LFP(2005)	0.031	0.021	0.045	0.086**	0.059**	0.097**	0.003	-0.011
	(0.019)	(0.034)	(0.029)	(0.041)	(0.027)	(0.038)	(0.021)	(0.023)
LFP(2006)	0.018	0.021	0.050*	0.067**	0.062**	0.087**	0.002	-0.011
	(0.027)	(0.039)	(0.028)	(0.030)	(0.031)	(0.038)	(0.023)	(0.024)
LFP(2007)	0.057**	0.081**	0.076**	0.100**	0.100***	0.153***	0.032	0.040
	(0.029)	(0.039)	(0.037)	(0.042)	(0.033)	(0.046)	(0.032)	(0.035)
Obs	248	502	187	395	221	448	214	449

Note: This table displays the effect of the return-to-work program on labor force participation (LFP) for different subgroups. LFP is equal to one if annual earnings exceed the SGA threshold. All subgroups, except for DI recipients aged 50-61, are based on on the sample of individuals aged 18-49 at the date of DI award. The estimates are based on the FD model and the RD model with controls. All covariates are measured prior to disability award, and described in Table (3). Standard errors (in parentheses) are robust to heteroscedasticity. In the FD model (described in Table 4), the sample is restricted to individuals awarded DI benefits during the period December 2003 - January 2004. In the RD model (described in Table 5), the sample is restricted to individuals awarded DI benefits during the period November 2003 - February 2004. 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.

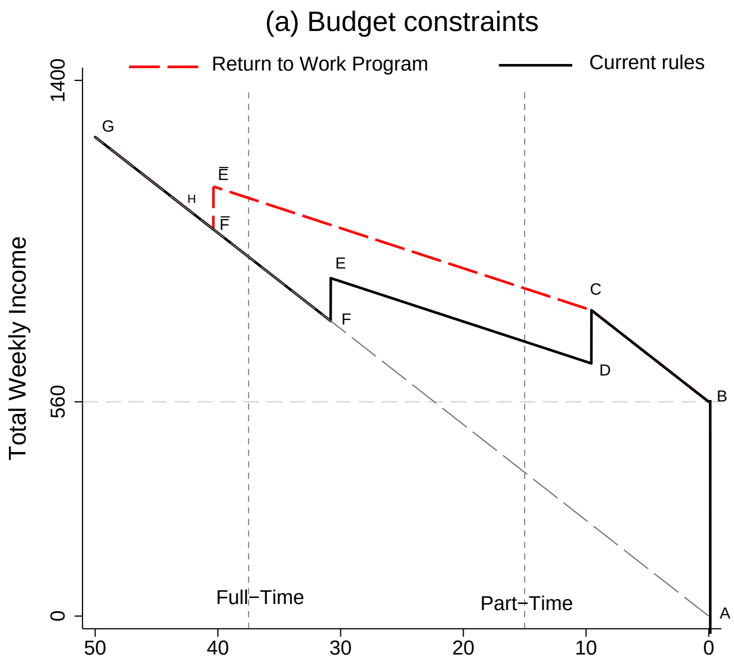
Table 11: AWARDED AND REJECTED APPLICANTS: LABOR FORCE PARTICIPATION AND CHARACTERISTICS

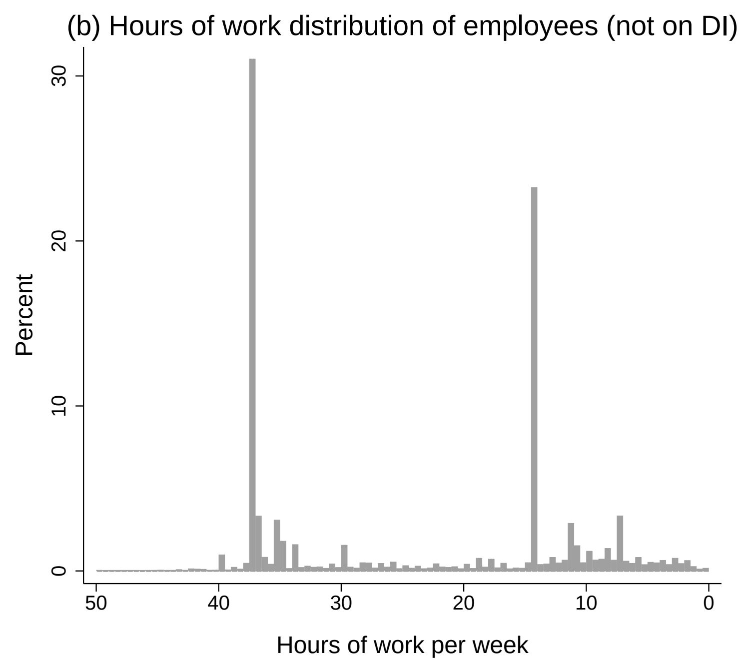
	LFP: Rejected Applicants	Mean difference in LFP: (Rejected Applicants - Awarded Applicants)			
Rejected/Awarded DI:	Nov. & Dec. 2003	Nov. & Dec. 2003		Jan. & Feb. 2003	
LFP(2005)	0.262	0.226***	0.288***	0.248***	0.300***
	(0.441)	(0.032)	(0.036)	(0.031)	(0.035)
LFP(2006)	0.307	0.257***	0.336***	0.288***	0.347***
	(0.462)	(0.034)	(0.037)	(0.033)	(0.036)
LFP(2007)	0.324	0.252***	0.332***	0.306***	0.369***
	(0.469)	(0.036)	(0.041)	(0.033)	(0.037)
Controls		NO	YES	NO	YES

	Characteristics: Rejected Applicants	Mean difference in characteristics: (Rejected Applicants - Awarded Applicants)	
Rejected/Awarded DI:	Nov. & Dec. 2003	Nov. & Dec. 2003	Jan. & Feb. 2003
Age at DI award	38.8	0.56	0.26
	(7.3)	(0.79)	(0.82)
Male	0.467	0.021	0.031
	(0.50)	(0.05)	(0.05)
Years of Schooling	10.1	-0.66**	-0.34
	(3.5)	(0.28)	(0.33)
Experience	10.1	-3.8***	-4.0***
	(8.2)	(0.82)	(0.87)
AIE	32949	-5330***	-5620**
	(14051)	(1364)	(1525)
Local unemployment rate	0.024	0.001**	0
	(0.007)	(0.001)	(0.001)
Local DI rate	0.096	-0.006***	-0.002
	(0.023)	(0.002)	(0.002)
Number of Children	1.013	0.08	0.05
	(1.26)	(0.11)	(0.12)
Musculosceletal system	0.4	0.177***	0.191***
	(.491)	(0.043)	(0.043)
Mental disorders	0.298	-0.148***	-0.107**
	(0.458)	(0.045)	(0.045)
Obs	225	445	440

Note: This table displays differences in labor force participation (LFP) and characteristics between rejected and awarded applicants. LFP is equal to one if annual earnings exceed the SGA threshold. The sample consists of individuals aged 18-49 at the time they were awarded/rejected DI. The first column shows mean and standard deviation (in parentheses) of LFP and characteristics for individuals whose DI application was rejcted in Nov. and Dec. 2003. Rejected applicants who successfully reapply during the following five years are excluded. The next two columns shows the mean difference in LFP (and characteristics) of the rejected applicants and individuals awarded DI in Nov. and Dec. 2003. The last two columns shows the mean difference in LFP (and characteristics) and standard errors (in paranthesis) of the rejected applicants and individuals awarded DI in Jan. and Feb. 2003. We show unconditional (no controls) and conditional (with controls) mean differences. When computing the conditional mean differences, we allow for a flexible functional form by including a large set of dummies for different values of the covariates (described in table 3) .

Figure 1: Work incentives

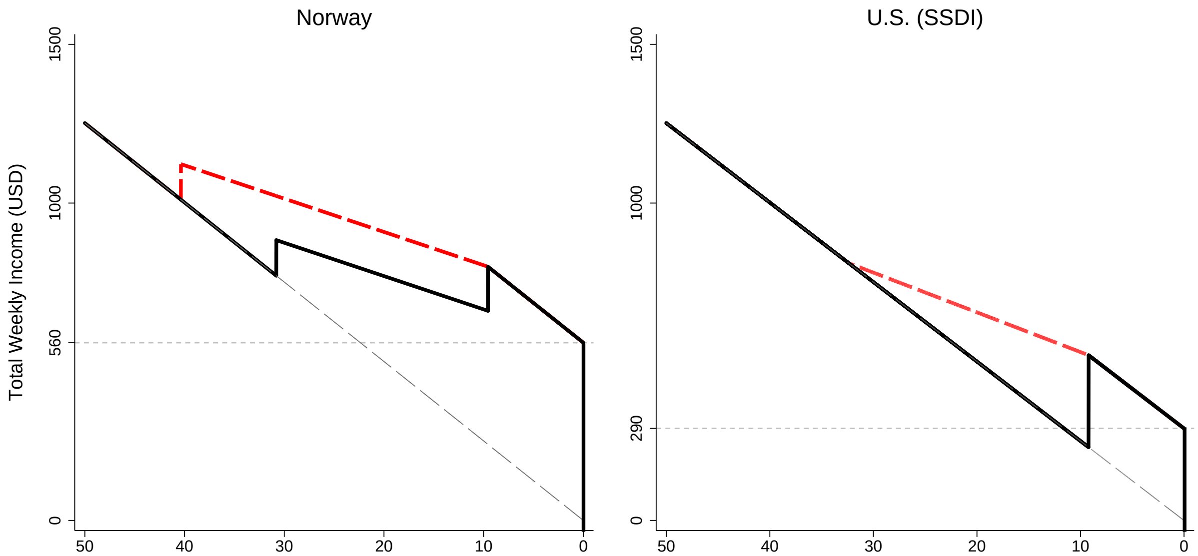




Note: Graph (a): The black solid lines represent the budget set for individuals under current rules and the dashed lines represent the budget constraint under the return-to-work program. We use the wage and benefit levels associated with a typical DI recipient. To compute the benefit levels, we use the average work history of recipients awarded DI during the period July 2003 to February 2004. We further set the wage equal to the average wage in a sample of workers who are not on DI but have similar observable characteristics as the DI recipients (see footnote 9). For simplicity, we disregard income taxation and dependent benefits: Total weekly income is thus equal to earnings and DI benefits.

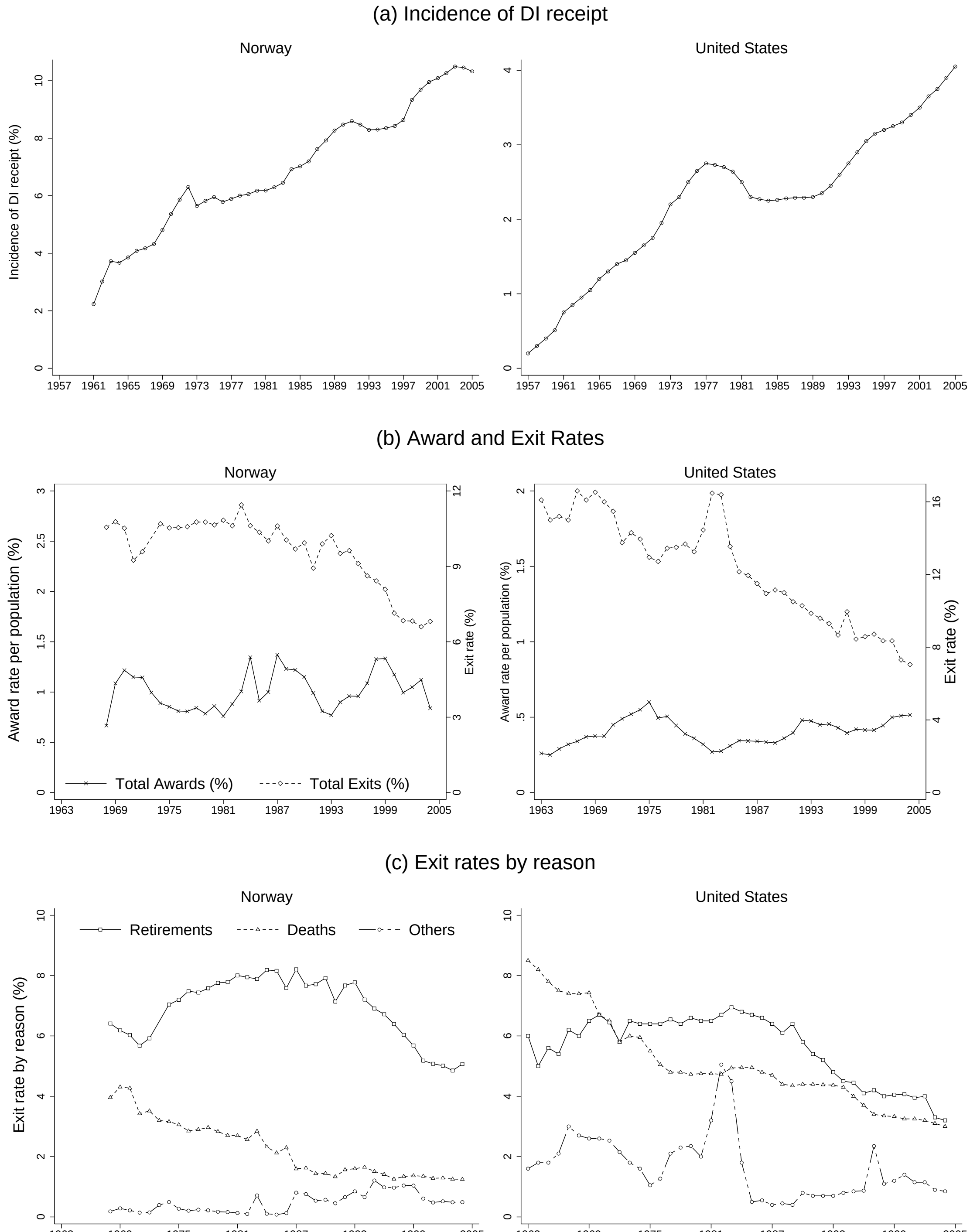
Graph (b): Displays the hours of work distribution of employees who are not receiving DI benefits. We use information from the Wage Statistics Survey in 2005. This survey covers all employees in the public sector. For employees in the private sector, the data is based on an annual stratified random sampling of all enterprises. Hours of work is defined as contractual number of working hours per week, excluding meal breaks.

Figure 3: Budget sets in Norway and the U.S.



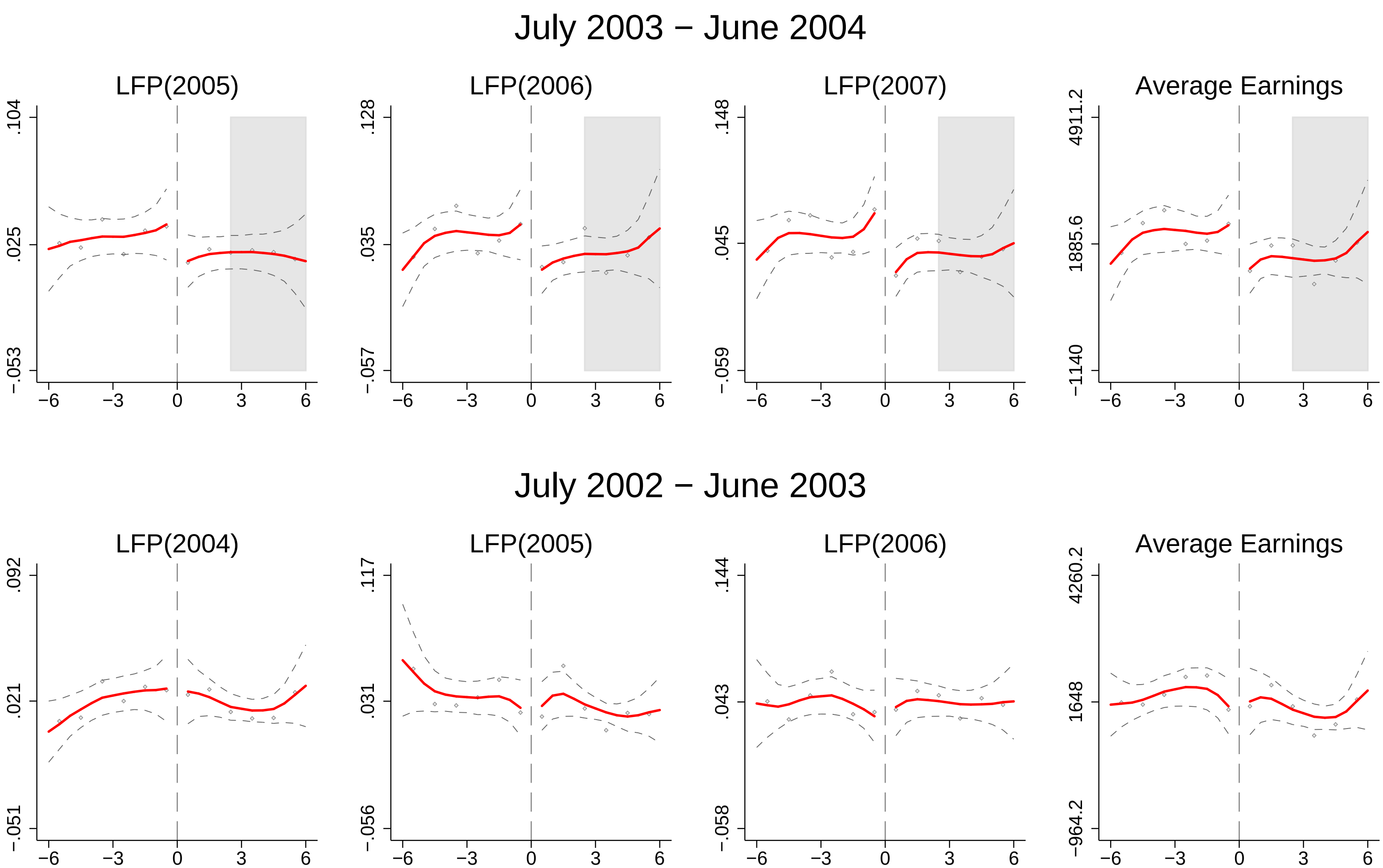
Note: The black solid lines represent the budget set for individuals under current rules, and the dashed lines represent the return-to-work program and the $1 for $2 offset budget constraints. In each budget set, we use the wage and benefit levels associated with a typical Norwegian DI recipient. To compute the benefit levels, we use the average work history of recipients awarded DI during the period November 2003 to February 2004. We further set the wage equal to the average wage in the sample of workers who are not on DI but have similar observable characteristics as the DI recipients (see footnote 9). Total weekly income equals earnings and DI benefits. For simplicity, we ignore income taxation, dependent benefits, and health insurance. The SSDI budget set under the $1 for $2 offset is constructed by calculating the AIE and subsequently the corresponding PIA-amount, according to the standard SSDI formula.

Figure 2: DI Trends in Norway and the U.S.



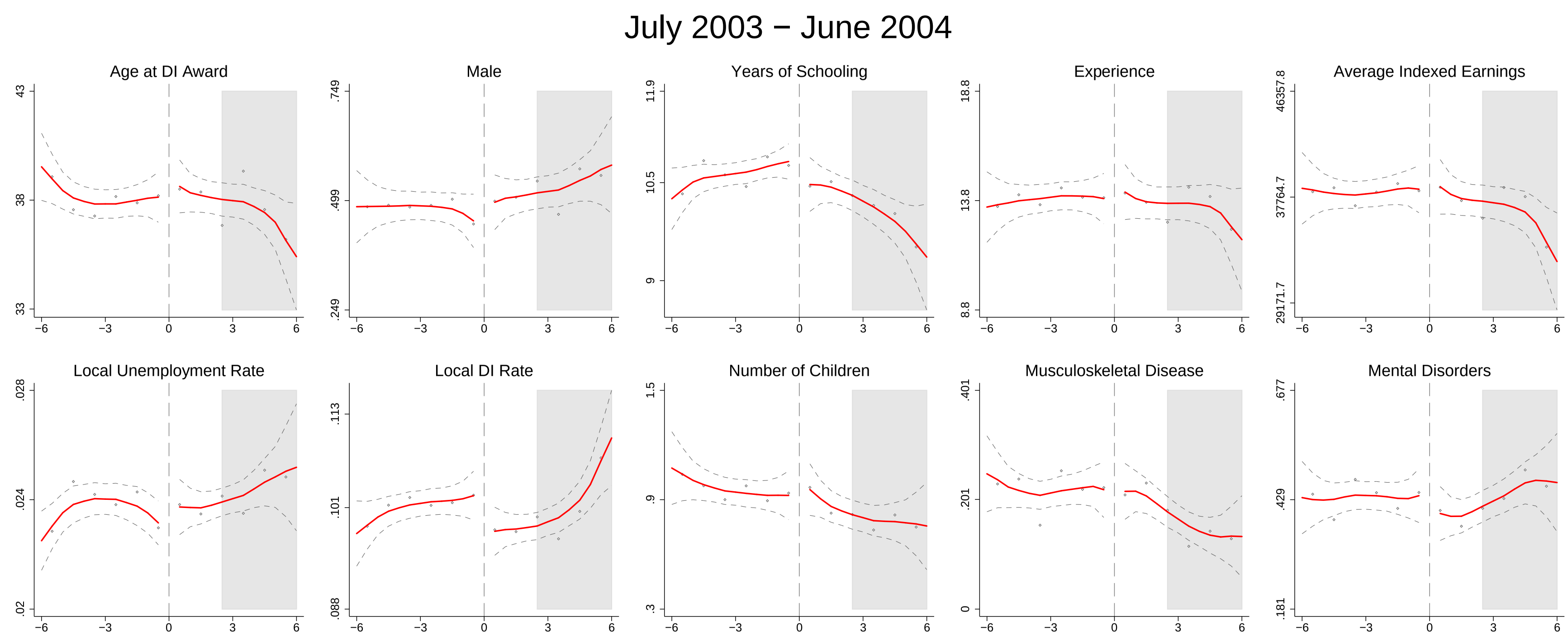
Note: The U.S. trends are based on Autor and Duggan (2006), while the Norwegian trends are collected from various issues of the Social Security Annual Statistical Supplement. Panel (a): Incidence of DI recipiency, defined as the percent receiving DI benefits in the adult population (aged 18 - 66 years 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 4: Labour Market Outcomes by Award Date



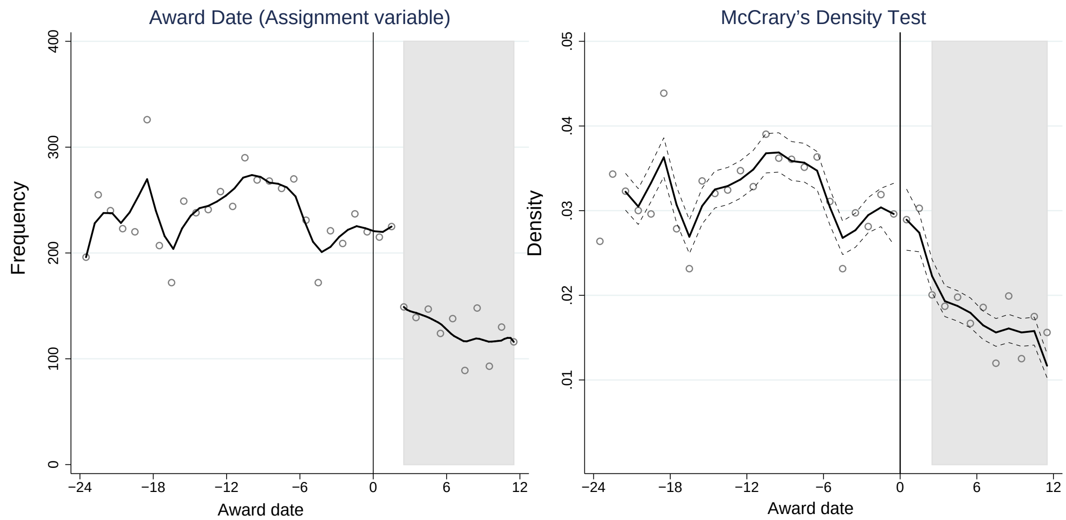
Note: The figure displays labor force participation (LFP) and mean earnings by month of DI award. LFP is equal to one if annual earnings exceed the SGA threshold. Average earnings is the mean earnings over the period 2005-2007. The top panel consist of individuals aged 18-49 (at the date of DI award), 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: January 2004 in top panel; January 2003 in bottom panel). The dashed lines represent the 95 percent confidence interval of the estimated monthly means. The grey 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: Characteristics by Award Date



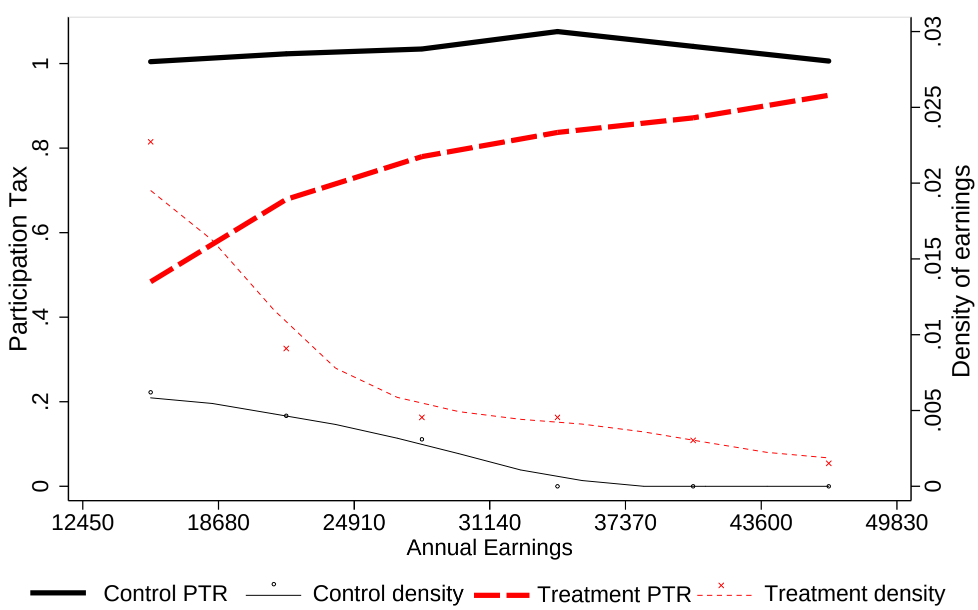
Note: The figure displays individual characteristics by month of DI award. All individual characteristics are measured prior to disability award, and described in Table 3. The sample consist of individuals aged 18-49 (at the date of DI award), who were awarded DI benefits during the period July 2003 to June 2004. 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, January 2004 (denoted 0). The dashed lines represent the 95 percent confidence interval of the estimated monthly means. The grey 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 $\pm.5$ of the respective variable.

Figure 6: Distribution of the Assignment Variable and the McCrary Test



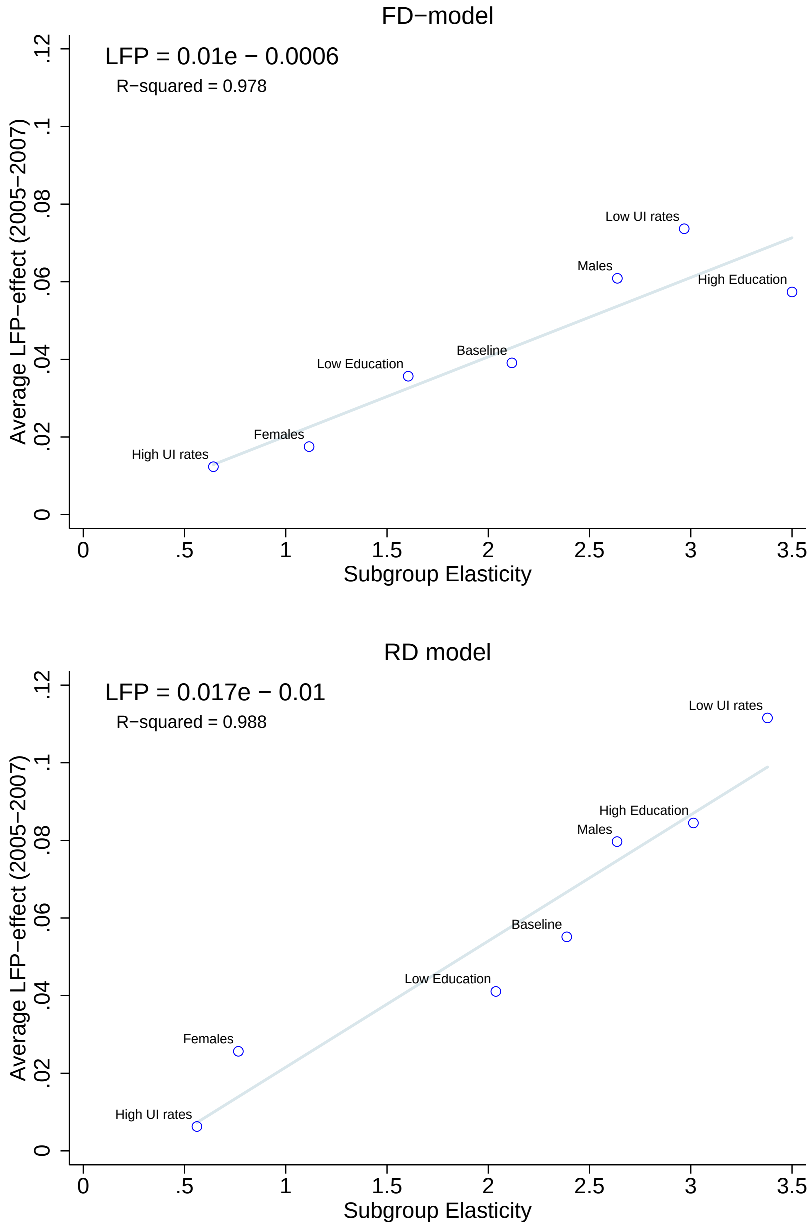
Note: The sample consist of individuals aged 18-49 (at the date of DI award), who were awarded DI benefits during the period January 2002 to December 2004. The left panel shows the frequency of DI recipients by month of DI award: It plots the unrestricted monthly means, and the estimated monthly means from a local linear regression applied to each side of the cut-off date for eligibility to temporary DI, March 2004. The right panel shows the density of DI recipients by month of DI award: It plots the monthly density, and then smooth the histogram using a local linear regression (with 2 months of bandwidth) applied to each side of the cut-off date for eligibility to the return-to-work program, January 2004. The dashed lines represent the 95 percent confidence interval of the estimated monthly means. In the x-axis, 0 represents January 2004. The grey shaded areas mark the months in which individuals awarded DI were affected by a potentially confounding policy change.

Figure 7: Earnings Densities and Participation Tax Rates



Note: This figure displays the the density of 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 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 bandwidth of 0.5 SGA. The dots represent the unrestricted density in each bin, while the thin lines represent the estimated density from local linear regressions.

Figure 8: Subsample Elasticities



Note: This figure plots the effect of the return-to-work program on labor force participation (LFP) against the absolute value of the aggregate labor force participation elasticities: For each subgroup, we display the average effect on LFP and the average elasticity, over the period 2005-2007. The subgroup estimates on LFP are reported in Table 10. The upper panel uses the FD model with covariates. The lower panel uses the RD model with covariates. The elasiticty and its components are computed separately for each subgroup. The participation tax rate (PTR) is given by equation (5), and incorporates the DI system, income taxation, and dependent benefits. The weighted sum of the differences in PTR between the treatment and the control group (ΔPTR) is given by equation (6). The elasticity is given by equation (4). The bases $\overline{PTR}$ and $\overline{LFP}$ are the weighted averages of PTR and LFP in the treatment and the control group. Each panel displays the regression line, coefficients and R-squared from a regression of average effect on LFP on the average elasticity and an intercept.