WORKING PAPERS IN ECONOMICS

No. 02/18

RAGNAR HJELLSET ALNE

ECONOMIC INCENTIVES, DISABILITY INSURANCE AND LABOR SUPPLY

Department of Economics
UNIVERSITY OF BERGEN

Economic incentives, disability insurance and labor supply

By Ragnar Hjellset Alne *

June, 2018

Abstract

Using a difference-in-difference model on full population data, I estimate the labor market response to a 2015 Norwegian disability insurance (DI) reform. The reform introduced an incentive program to encourage DI beneficiaries to increase their labor supply, and I find that the program significantly increased the average working hours and modestly affected the labor market participation of DI beneficiaries. There is significant heterogeneity in the estimated effects; young beneficiaries respond positively along the extensive and the intensive labor supply margins. The analysis accentuates the importance of analyzing both labor supply margins when evaluating the effects of DI reforms.

JEL Classification: D60;H53;I38;J08;J22

Keywords: Disability insurance reform, labor supply, economic incentives, difference-in-difference, labor economics

*Contact: University of Bergen, Department of Economics, P.O. 7802, 5020 Bergen, Norway. E-mail: ragnar.alne@uib.no. The Norwegian Labor and Welfare Administration provided the data used in the analysis, but the results and conclusions in this article should be attributed to the author only. I thank Espen Bratberg, Alf Erling Risa, Kjell Vaage, Julian Vedeler Johnsen, Otto Sevaldson Lillebø, Mirjam Wentzel, Nina Serdarevic, Eirik Strømland and Erling Risa for helpful comments. I am also grateful for comments received while presenting the paper at the 2017 NHH PhD Workshop, the Department of economics at the University of Bergen, the Ministry of Labor and Social Affairs, Western Norway University of Applied Sciences, the TREfF-2 Workshop at Solstrand, as well as the Norwegian Labor and Welfare Administration.

1 Introduction

The US government implemented The Ticket to Work and Work Incentives Improvement Act in 2002 to provide disability insurance (DI) beneficiaries with economic incentives to return to the labor force. The implementation of the act was an attempt to reduce the DI recipiency rate, which has increased with time in the US. High recipiency rates are also common in other countries; figure 1 shows that 6% of the working age population in the Organisation for Economic Co-operation and Development (OECD) countries claimed DI benefits in 2008. The fraction of the working age population who claims DI benefits also increased in many countries since the mid-1990's, and these countries will have to reverse the trend of an increasing recipiency rate to maintain the long-term viability of their DI programs.

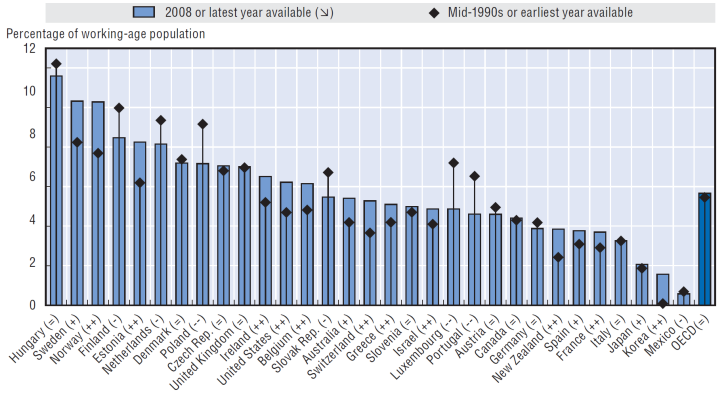

Figure 1: Development in the number of DI beneficiaries as a fraction of the working age population (OECD (2009)).

The growth in the DI recipiency rate has to some extent been predictable, because the average age in the workforce has increased. Older workers are often eligible to apply for a DI benefit, and they are also more likely to have a health condition that qualifies them for the insurance claim. Duggan and Imberman (2009) estimate that aging of the workforce accounts for 15% of the increase in the DI recipiency rate of men in the US, and for 4% of the increase in the recipiency rate of women. Autor and Duggan (2003) highlight the increased income inequality as another factor contributing to the growth in DI claims in the US, where benefits are indexed to the average wage level. This indexation has increased the real replacement rate in the DI program for low-wage workers, and therefore raised the attractiveness of a benefit claim relative to work for this group. The increased female labor force participation has also caused growth in the scale of the DI programs, by increasing the proportion of females that are eligible to apply for DI benefits. In most countries, the increased scale has increased the total costs of the DI

programs, and to avert a fiscal crisis in the US DI program the Bipartisan Budget Act of 2015 temporarily reallocated a larger share of the payroll taxes to the DI trust fund (Hill et al. (2016)). Currently, the US spend approximately 1.4% of its gross domestic product (GDP) on the DI program, while the average OECD country spends around 2.1% (OECD (2018)). However, there is significant variation in spending between countries, and countries such as Sweden and Denmark spend as much as 4-5% of their GDP on the DI program (OECD (2018)).

The problems of high costs and increasing recipiency rates have brought about numerous DI reforms and programs. These reforms can be sorted into three categories, where the first of these aims to reduce the entry rates through lower income replacement rates or stricter screenings into the DI program. Sweden and the Netherlands successfully reduced their DI benefit levels in the 1980s and 1990s; however, the approach of changing the DI benefit level to reduce the entry rates has its limitations. A DI program needs to provide a certain level of insurance against the loss of the ability to earn an income, and this insurance aspect needs to balance with the economic disincentive effects of the benefit payments. After deciding a certain level of insurance compensation, countries need to focus on different types of reform to minimize the disincentive effects of the DI program. The trade-off problem between economic incentives and insurance has led to a second type of reform that focuses on increasing the exit rate from the DI program through re-evaluations of the disability level, job counseling, or return to work programs. These exit-rate reforms have generally had modest effects; consequently, some countries have taken a third approach, using economic incentives to encourage DI beneficiaries to utilize their potential labor supply. The economic incentive approach to DI reform is more dynamic than the other two types of reform, because it encourages beneficiaries to work more if the health condition permits a higher labor supply without altering the insurance aspect of the DI program.

The implementation of The Ticket to Work and Work Incentives Improvement Act in the US was an early attempt at the economic incentive approach to DI reform. The act offered career counseling and job placement, but it also allowed an expedited reinstatement of DI benefits if a beneficiary's attempt to return to the labor force failed. The Ticket to Work and Work Incentives Improvement Act therefore reduced the economic risks associated with a return to the labor force, and Autor and Duggan (2006) found a moderate labor market response after the implementation the reform. The 2015 Norwegian DI reform I analyze here implemented some of the features from the Ticket to Work and Work Incentives Improvement Act; section 3.1 presents these features in detail. The UK also applied the economic incentive approach to DI reform in a program called Pathways to Work, which subsidized low-income beneficiaries with a small weekly payment if

they attempted a return to the workforce. Blyth (2006) found an eight percentage points increase in the exit rate from the DI program after the Pathways to Work program was implemented.

Norway furthered the incentive approach in a full population DI reform carried out in 2015, which aimed to mobilize the potential labor supply of DI beneficiaries through the use of economic incentives. I estimate the average treatment effects (ATE) of this reform in a difference-in-difference (DD) framework, and rich data enables me to estimate the causal effects of the 2015 reform. With access to data on reported working hours, I estimate both the intensive and the extensive margin labor market response, which measures the average change in the labor force participation and the average change in working hours conditional on employment. I also perform a heterogeneity analysis to investigate variation in the labor market responses across sub-groups that are categorized by age, gender and disability level. The working hours analysis contributes to the DI literature, which has focused primarily on the extensive margin labor supply so far.

The main results from the analysis indicate that young individuals and men respond most strongly to the changes in the incentive structure at the extensive margin. Therefore, the extensive margin estimates support prior results in the DI literature. I also find a decreasing total labor market response with respect to age in the working hours analysis, where partial DI beneficiaries drives this decreasing age effect. There is a clear, positive increase in the working hours after the reform, and the estimated working hour effects are consistently higher for men and 100% beneficiaries than for women and partial beneficiaries. Interestingly, full DI beneficiaries had the strongest average treatment response to the changes in the incentive structure at the intensive margin. This suggests that an economic program that continually incentivizes DI beneficiaries to utilize their potential labor supply, can mobilize individuals of poor health when the health condition permits a higher labor supply. A sound incentive program will therefore maximize the labor supply of DI beneficiaries over time, and potentially improve the long-term viability of the DI program through this dynamic approach.

In section 2, I present the research that is most relevant and related to this paper. Section 3 introduces the Norwegian social insurance program and important information about the economic incentive program in the 2015 DI reform. Section 4 explains the empirical strategy and the data used in the analysis, while section 5 presents the summary statistics. The results from the empirical analysis are found in section 6, and I summarize the results and discuss the contributions of this paper in section 7.

2 Related research

The literature that focuses on the incentive effects of DI on labor market outcomes dates back to Parsons (1980). He estimated that DI could explain explain most of the increase in the labor force non-participation rate among older males in the US between 1948 and 1976. Haveman and Wolfe (1984) found significantly lower disincentive effects and estimated that the DI program could explain approximately 20% of the increase in the male non-participation rate in the US between 1968 and 1978. The disincentive effects of the DI program has later been discussed in numerous papers as a trade-off between the economic incentives and the insurance aspect of a DI program; see e.g., Bound et al. (2004), Benıtez-Silva et al. (2006) and Bound et al. (2010).

Other studies have estimated the effect of receiving or losing the right to claim DI benefits. Moore (2015) found that a large fraction of DI beneficiaries with addiction disorders find jobs after losing eligibility for a DI claim. Young individuals are most likely to return to the labor force after losing the eligibility; Moore (2015) concluded that evaluations of the work capacity and the health condition at the time an individual applies for a DI benefit might not be a good predictor of future work capacity. Consequently, it is important to focus on dynamic effects when optimizing the incentive structure in a DI program. French and Song (2014) similarly found that the award of benefits reduce the labor force participation of young DI applicants.

A number of studies have focused on the effects of changing the DI benefit level. Borghans et al. (2014) found that the earnings of existing beneficiaries increased after a Dutch reform reduced the DI benefit level. Gruber (2000) applied a DD strategy and found that an increased benefit level caused an increase in the non-employment in Canada. However, Campolieti (2004) could not find a similar response to a comparable Canadian DI reform, and argued that time-variation in the medical screening stringency can account for the discrepancy in the estimated effects between Gruber (2000) and Campolieti (2004).

Bound (1989) was the first to suggest that individuals that are rejected from the DI program form a natural control group for DI beneficiaries. He found that approximately one-third of rejected DI applicants work after being rejected from the DI program; this is a good upper-bound estimate for the fraction of DI beneficiaries that would be working in the absence of a DI program, since rejected applicants are of better health and more capable of working than those accepted into the DI program. Using Bound's (1989) idea, Wachter et al. (2011) found that young beneficiaries have a significant unused potential labor supply, where the highest potential for an increase lays with individuals that have disorders such as depression or back pain. The fraction of beneficiaries with these

disorders has increased with time in many countries, and this could signal an increased potential labor supply in the DI population in general. Both Bound (1989) and Wachter et al. (2011) acknowledged that most DI beneficiaries would have a weak labor force attachment in the absence of a DI program, and Maestas et al. (2013) argued the labor supply of individuals at the margin of program entry would increase in the absence of a DI program. Low and Pistaferri (2015) further argued that false rejections from the DI program are more common than the number of false acceptances to the DI program.

The DI literature most closely related to this article focuses on the use of economic incentives to encourage DI beneficiaries to increase their labor supply, without altering the insurance aspect of the DI program. Using simulations, Moffitt (1999) showed that welfare reforms intended to incentivize labor supply increases could reduce the labor supply in the incentivized group. The Moffitt (1999) paper accentuates how difficult it is to create an efficient incentive structure; the DI reform analyzed here targets this problem.

Aakvik et al. (2001) analyzed the successful implementation of a 1997 Norwegian reform, which doubled the level of income a DI beneficiary could earn while claiming full DI benefits. This increase of the significant gainful activity (SGA) threshold increased both the intensive and the extensive margin labor supply of DI beneficiaries, and in Norway the SGA threshold has remained unchanged since this 1997 reform. However, the 2015 reform analyzed here alters the SGA threshold; section 3.1 presents the average pre- and post-reform budget constraints in detail. The findings in Aakvik et al. (2001) is backed by Campolieti and Riddell (2012), who also found that an increased SGA threshold in the Canadian DI positively affected the labor force participation. Interestingly, Campolieti and Riddell (2012) found the largest labor market response among women, who increased the labor force participation by 7.9 to 9.5 percentage points.

In 1999, the US Congress mandated a project to test the effects of a gradual offset of DI benefits with an increasing salary. The implemented project reduced the DI benefits with \$1 for every \$2 a DI beneficiary earned above a defined SGA threshold. This reduction model replaced a sharp discontinuity in the budget constraint, which significantly reduced the DI benefits when an income marginally exceeded the SGA threshold. Weathers and Hemmeter (2011) found that the implementation of the reduction model increased the probability of having an income above the SGA threshold, but the gradual offset did not affect labor force participation. I extend the analysis of Weathers and Hemmeter (2011) by also analyzing the intensive margin labor market response, measured by the change in reported working hours.

Kostøl and Mogstad (2014) analyzed a DI reform, which was the precursor to the reform I analyze here. The pilot reform allowed some DI beneficiaries to keep a larger

fraction of their income after increasing their labor supply, and Kostøl and Mogstad (2014) estimated the local average treatment effects (LATE) of the reform on the labor supply of full DI beneficiaries in a regression discontinuity (RD) framework. They found the probability of having an income above the old SGA threshold increased by nine percentage points, and men and young DI beneficiaries responded most strongly to the economic incentives. Kostøl and Mogstad (2014) concluded that DI beneficiaries have a significant unused potential labor supply, which economic incentives have the potential to mobilize. Koning and van Sonsbeek (2017) performed a similar analysis to Kostøl and Mogstad (2014) for partial DI beneficiaries in the Dutch DI program. While Koning and van Sonsbeek (2017) found no evidence that partial beneficiaries had a larger response than full beneficiaries, both Kostøl and Mogstad (2014) and Koning and van Sonsbeek (2017) found the largest labor market response among young DI beneficiaries. I expand on the analysis in Kostøl and Mogstad (2014) and Koning and van Sonsbeek (2017) by analyzing both the intensive and the extensive margin labor supply responses of full and partial DI beneficiaries, but in a DD framework as opposed to an RD framework. I therefore estimate an ATE, as opposed to a LATE, because this is an important parameter for policy-makers.

3 Social insurance in Norway

Established by the implementation of the National Insurance Act in 1967, the Norwegian social security program is a universal tax-funded public insurance program managed by the Norwegian Labor and Welfare Administration. The social security provides insurance such as retirement pension and unemployment insurance, and health-related insurance such as sick pay, temporary DI, and permanent DI. All the insurance provided by the social security are universal for residents, although prior income is a requirement for eligibility to apply for some of them. The most relevant programs for this analysis are those that insure workers against a loss of income due to injury or illness.

Sick pay compensates eligible individuals 100% of lost earnings the first year after they exit the labor force due to sickness or injury. After receiving sick pay for one year, a four-year temporary DI program is available for eligible individuals.[1] The primary purpose of the temporary DI program is to evaluate and improve a beneficiary's ability to work, with an overall goal of returning him or her to the labor force. If the temporary DI program fails to return an individual to the labor force, a permanent DI program is available for individuals in poor health. Most beneficiaries enter the permanent DI program from the temporary one, and I analyze the permanent DI program in this paper.

[1]The temporary DI program was four years when the 2015 DI reform was implemented, but after 1 January 2018 the temporary DI program is reduced to a three year program.

In 1967, the number of permanent DI beneficiaries was one-third of the number in 2015. Beneficiaries were divided equally between men and women, and figure 2 shows that the growth in the number of female beneficiaries has been more rapid than that of men. Figure 3 shows that the DI recipiency rate doubled after 1967, and the two figures therefore indicate that the number of DI beneficiaries has grown faster than the DI recipiency rate. Population growth therefore partially explains the increased number of DI beneficiaries, and because of this growth Norway spends approximately 3.7% of its GDP on the DI program; this is significantly more than the OECD average of 2.1% (OECD (2018)).[2]

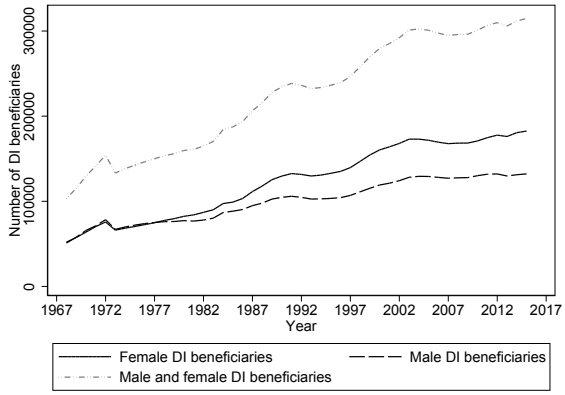

Figure 2: Number of DI beneficiaries from 1967 until 2015. Data is provided by the Norwegian Labor and Welfare Administration.

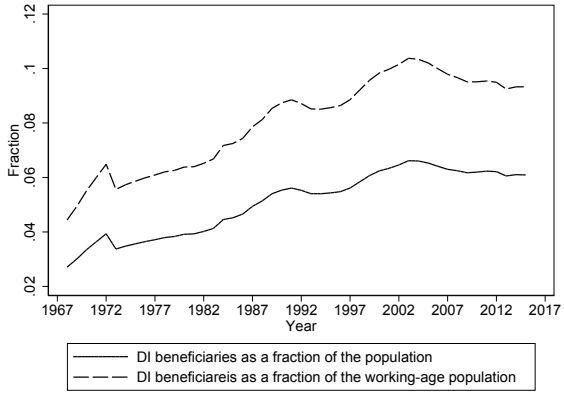

Figure 3: DI beneficiaries as a fraction of the population. Working age is defined as 18–67 years. Data is provided by the Norwegian Labor and Welfare Administration.

It is evaluations of medical doctors that qualify individuals for a health-related insurance claim in Norway. A primary care doctor writes a medical declaration for an individual making a health-related insurance claim. Disability examiners and medical doctors, employed by the Norwegian Labor and Welfare Administration, evaluates this medical declaration to verify the validity of the claim. The disability examiners and the medical doctors therefore have a central role in the application process, and they are crucially important to the validation of health related insurance claims in Norway.[3]

Eligibility for the permanent DI program requires a 50% reduction in the ability to earn an income, but a 30% reduction is sufficient when the debilitating injury or illness is work-related. 82% of all DI beneficiaries are evaluated to have fully lost their ability to

[2]The Norwegian spending on the DI program is hardly comparable with the spending of other countries, since the revenues from the oil and gas industry gives Norway a deflated spending measure due to the high GDP.

[3]Dahl et al. (2014) presents the application process to the Norwegian DI program in great detail.

earn an income, and this makes them eligible for a 100% DI claim.[4] As many as 12.5% of all beneficiaries claim 40-60% DI benefits, because they have partially lost their ability to earn an income. The DI program's formal replacement rate is 66% of an estimated income that limits the minimum and the maximum benefit level.[5][6] Prior to the 2015 DI reform, the estimated income was calculated from projected retirement savings, as if the person had continued work until the age of 67. After the 2015 DI reform, the estimated income is the applicant's highest average income in three of the five years prior to the point of disability, limited by the minimum and maximum estimated income levels in the DI program.

In 2011, the pension system underwent a major reform, which was important to the 2015 DI reform. The pension reform implemented economic incentives for individuals to stay in the labor force in old age. The new system allows individuals to save for their pension until the age of 75, and the 2011 reform separated the decision to retire from the decision to leave the labor force. Retirees can therefore claim actuarially fair retirement benefits from the age of 62, while simultaneously working past the age of retirement. An early withdrawal of pension benefits lowers the payments through an age-adjustment, and the retirement system also make a longevity adjustment of the benefits based on the average life-expectancy of each birth-cohort. However, the separation of the decision to retire from the decision to leave the labor force caused a discrepancy between the retirement system and the permanent DI program. This discrepancy provided some individuals with incentives to apply for acceptance into the permanent DI program towards the end of their working career; these perverse incentives had to be removed from the welfare system. The introduction of the 2011 retirement reform was therefore a catalyst to the introduction of the economic incentive program in the 2015 DI reform.

3.1 The 2015 DI reform

A central goal of the 2015 DI reform was to use economic incentives to continuously encourage DI beneficiaries to increase their labor supply. Prior to 2015, 100% DI beneficiaries could earn $11,846 per year and simultaneously claim full DI benefits. Exceeding this threshold resulted in approximately a 25% reduction in the disability level and the DI benefits concurrently.[7] Consequently, the pre-reform budget constraint was piecewise linear, with a sharp discontinuity at the SGA threshold; figure 4 shows a graphical approximation of the average pre- and post-reform budget constraints.

[4]The distribution is calculated for all observations of DI beneficiaries in 2014 and 2015.

[5]Minimum estimated income = $44,460, maximum estimated income = $71,136, minimum DI befits = $29,403, maximum DI befits = $46,950

[6]All conversions use the benefit levels and the exchange rate from 1 January 2015. The 7.4605 USDNOK exchange rate is low, relative to the average exchange rate in the period after the DI reform.

[7]Appendix B explains the calculation of the reduction in the DI benefits and DI benefit level.

Figure 4 shows that the 2015 reform modified the old budget constraint, by introducing a reduction model above a new \$8,042 SGA threshold for all existing DI beneficiaries that claim full benefits.[8] Above the new SGA threshold, benefits are gradually reduced against an increasing salary by the DI program's real rate of income replacement. The 2015 reform therefore introduced a continuous and kinked budget constraint.

The 2015 reform also introduced other economic incentives. It awarded DI beneficiaries a permanent right to the DI benefit payments, as well as the level of disability. The reform also removed a re-evaluation of the disability-level, which took place when the yearly income exceeded the old \$11,846 SGA threshold. The award of a permanent right to the DI benefits, combined with the removal of the re-evaluation of the disability level, is similar to the expedited reinstatement of the DI benefits in The Ticket to Work and Work Incentives Improvement Act in the US. It eliminates the economic risks that were associated with entering the labor force prior to the 2015 reform, and it replaces a five-year trial period in the labor market to which beneficiaries were entitled prior to 2015.

The 2015 reform also removed a one-year waiting-period after the award of DI benefits, before beneficiaries could enter the labor force. This period of isolation from the labor force could result in degradation of skills, and reduce the attractiveness of DI beneficiaries in the labor market. In addition, the 2015 reform changed the taxation of DI benefits from a pension tax to an income tax. This tax change results in increased transparency in transfers between the welfare system and the labor market. The reform therefore removed some of the opaqueness to the tax-system, which might have been a barrier to entering the labor force prior to the reform. To counteract the effective tax-increase, the Norwegian Labor and Welfare Administration increased the pre-tax benefits, with an overall goal of improving the labor market outcomes without affecting the insurance aspect of the DI program.

[8]The \$8,042 SGA threshold is reduced to \$4,742 in 2019. New beneficiaries that claim full benefits after the 2015 reform have a \$4,742 SGA threshold from the date of acceptance into the DI program.

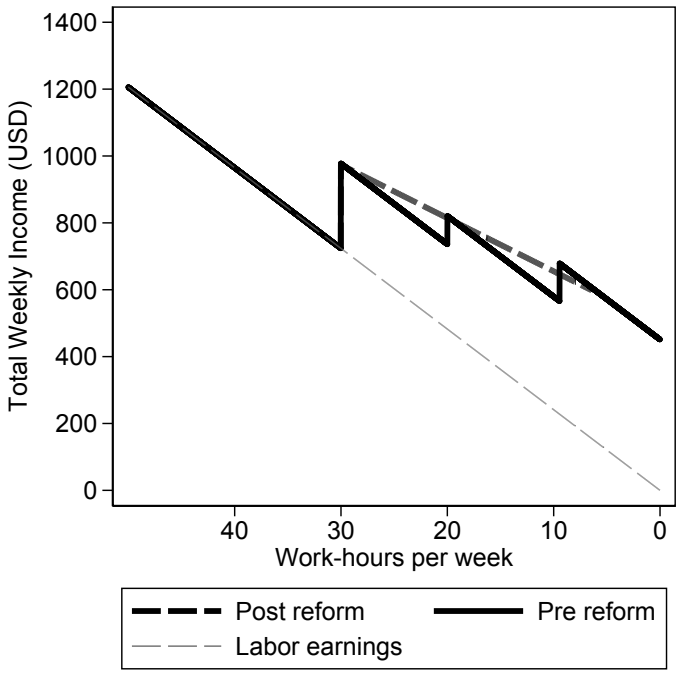

Figure 4: Graphical approximation to the average pre- and post-reform budget restriction for 100% DI beneficiaries, as a function of weekly working hours. I have assumed an average hourly salary of approximately \$24, an average weekly benefit payment of \$451, and a real replacement rate of 66% to make a reasonable approximation to the average individual who claims full DI benefits in Norway. Appendix B provides a more exact representation of an individual's pre- and post-reform budget constraints, as a function of yearly earnings.

Because the DI reform leaves the net benefit level unchanged, it is unlikely that the DI reform will have a large effect on the extensive margin labor supply of beneficiaries. However, the reform removed the economic risks associated with entering the labor force, which might encourage individuals to utilize their work capacity if the health situation improves after some time in the DI program. The economic incentives might also encourage individuals with variable health conditions to work in periods when the health situation improves, because they are entitled to restart a benefit claim if an attempt to increase the labor supply fails after the reform.

It is likely that the labor market demand for individuals with variable or poor health conditions is of importance to the extensive labor market response following the DI reform. Employers might be skeptical about hiring beneficiaries that enter and exit the labor market according to changes in their health. However, it might be easier for DI beneficiaries to change the number of working hours at an existing employer. I therefore expect the DI reform will primarily affect the intensive margin labor supply, where the most affected individuals are those who adapted their labor supply around the first

discontinuity in the old budget constraint prior to the reform. The removal of the re-evaluation of the disability level at the old SGA threshold should positively affect the intensive margin labor supply of DI beneficiaries. However, the reduction model will increase the marginal tax rate for incomes above the new SGA threshold; this, in turn, should negatively affect the working hours of DI beneficiaries. The net change in the working hours after the reform is therefore unclear— it might be positive, negative or unchanged. If the 2015 reform increased the labor supply, DI beneficiaries had an unused potential labor supply prior to 2015, which is in demand and mobilized by the economic incentive program after the reform. The 2015 reform is likely to reduce the labor supply if DI beneficiaries did not have an unused potential labor supply prior to the DI reform. The reduction model will then function as an increased marginal tax rate and reduce the economic incentives to work after the reform. The positive and negative effects from the economic incentives program might also balance out evenly, and leave the average labor supply of DI beneficiaries unchanged after the reform.

Even if it is difficult to predict an overall effect of the DI reform, it is likely that the old program had a ratchet effect on the intensive margin labor supply of DI beneficiaries. This ratchet effect came from the discontinuity in the pre-reform budget constraint, which provided DI beneficiaries with few incentives to reveal their true potential labor supply. The ratchet effect is clearly visualized in the yearly income distributions of 100% DI beneficiaries in figure 5, where the 2013 and 2014 histograms show that many individuals adapted their labor supply below the pre-reform $11,846 SGA threshold. However, there is no adaptation around the new $8,042 SGA threshold prior to the 2015 DI reform. In 2015, when DI beneficiaries no longer have incentives to adapt below $11,846, some beneficiaries adapt their yearly labor supply around $8,042, where the reduction model is introduced. DI beneficiaries pay a high marginal tax rate above this $8,042 SGA threshold in 2015, which leaves them with reduced incentives for a higher labor supply. However, the right tail in the 2015 income distribution is thicker than what it was prior to the DI reform, and it seems clear that the probability of earning an income above the old SGA threshold has increased after the DI reform. It is therefore likely that a LATE estimate, that estimates the change in the probability of earning an income above the old SGA threshold, would find a positive effect of the DI reform. However, I find it more relevant to estimate the effect of the 2015 reform on the average labor supply in the DI population. The histograms in figure 5 seem to indicate that the reduction model has made the incentives smoother and more continuous for the DI beneficiaries. The reform has led some beneficiaries to reduce their labor supply to around the kink in the new budget constraint, while some DI beneficiaries increased their labor supply above the old SGA threshold after the reform. An empirical evaluation is necessary to determine the

average overall effect of the reform.

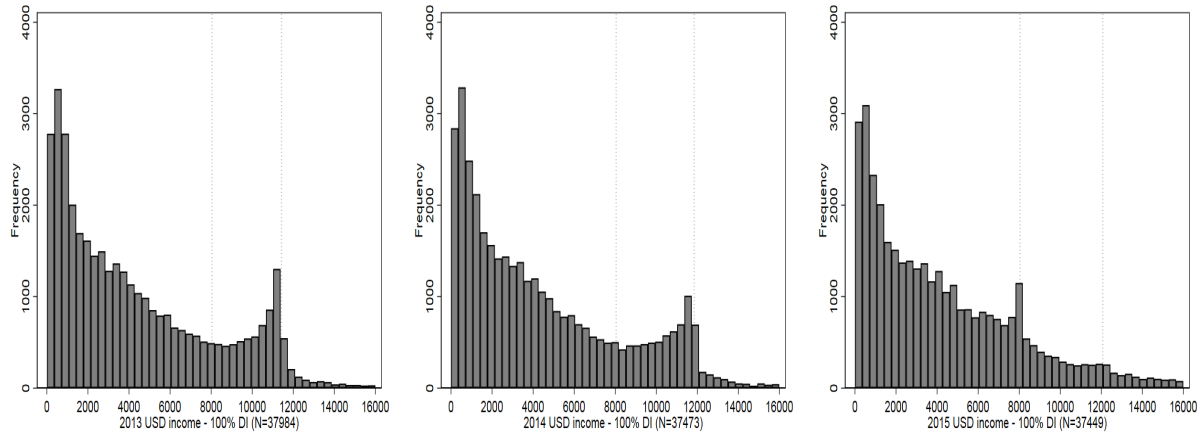

Figure 5: Distribution of labor income (>$0) for full DI beneficiaries from 2013-2015. N indicates the total number of observations in the histogram.

4 Empirical strategy

I use a difference-in-difference (DD) model to capture the causal labor market response after the 2015 DI reform. The DD model estimates an average treatment effect (ATE), which might not be a good estimate for the treatment response of distinct subgroups affected by the DI reform. I address this problem by performing a thorough heterogeneity analysis across groups categorized by age, gender and disability level. The analysis should capture potential heterogeneous labor market responses after the DI reform, and thereby address the weakness of an ATE estimate.

In an ideal scenario, the 2015 DI reform would have randomized beneficiaries into the new program as a treatment group and kept a fraction of the beneficiaries in the old system as a control group. However, the 2015 reform transferred all existing beneficiaries from the old to the new DI system, which makes it a challenge to find a natural control group for the analysis. The main problem for the analysis is therefore to find a control group suitable for the estimation of the causal labor market response after the DI reform.

Bound (1989) first suggested that rejected DI applicants form a natural control group for DI beneficiaries. His idea partly motivates my choice of control group in this analysis. Rejected DI applicants have similar traits to DI beneficiaries, such as a high average age, a low labor supply, and a similar gender balance. More importantly, they are, on average, in poor health; this is the central reason to consider them as a control group in the analysis. The descriptive statistics in section 5 provide support for rejected DI applicants as the control group, which controls for unobserved heterogeneity through the fixed group difference in the DD model.

The inclusion of a trend adjustment in the DD model loosens the parallel trends assumption, which is the main identifying assumption in the DD framework. The trend adjustment removes a potential source of bias in the analysis, which tests for heterogeneous responses in numerous sub-groups. However, I report both trend-adjusted and non-trend-adjusted estimates for all regressions in section 6, and the main conclusions in the analysis are unaffected by the inclusion of this trend-adjustment. The DD model in the analysis is:

$$Y_{it} = \alpha_i + \alpha_1 post_t + \beta(treat_i * post_t) + \sum_{j=jan14}^{dec15} \alpha_j T_{jt} + \alpha_2 t + \alpha_3(treat_i * t) + \epsilon_{it}$$

Subscript i specifies the individual and t represents time measured in month and year. Y_{it} is a log-transformed working hours variable in the intensive margin labor market analysis, which estimates the percentage point change in working hours after the DI reform. I log-transformed the working hours variable, because the fraction of individuals that work zero hours is high among DI beneficiaries. Including the non-workers would not make an interesting intensive margin labor market response analysis.

In the extensive margin analysis, Y_{it} is a dummy variable that equals one if the labor supply exceeds an average of four working hours per week in the observation month. Prior to 2015, an individual had to work four hours per week to trigger monthly reporting from employers to the social security register; therefore, four working hours per week is the defining threshold of the extensive margin. I follow Blundell et al. (2013), who argue that it is best practice to use working hours and the probability of working to measure the intensive and extensive margin labor supply.

Post is a variable that equals one for observations made during the first twelve months following the DI reform and zero for observations made in the twelve months prior to the reform. *Treat* is a dummy variable that equals one for individuals accepted into the DI program prior to the 2015 reform. *Treat* equals zero for individuals rejected from the DI program between 2008 and 2013, who is not accepted into the DI program at appeal by December 2014. Therefore, it is strictly pre-reform characteristics that define the assignment to the treatment and the control group. β is an estimate of the average change in the labor supply of DI beneficiaries relative to rejected DI applicants, and it is an estimate of the causal labor market response the first year after the 2015 reform.

T is dummy variables per month per year that control for temporal changes to the labor supply that is common to all individuals. α_i is the individual fixed effects, and controls for fixed unobserved heterogeneity between individuals. There might be some time-varying heterogeneity coming from family, industry or place of residence, but these characteristics are assumed to be stable across the two groups in the analysis. The main

specification of the DD model does not include control variables, but the estimated effects are not sensitive to the inclusion of control variables. The trend-adjusted model includes a continuous time-variable t, as well as an interaction term between the *treat* dummy and the time-variable t. The interaction term captures potential linear differences in time-trends between the treatment and control group, and ϵ_{it} is the error term in the model.

All regressions are clustered at a municipality level to adjust the standard errors for heteroscedasticity and serial correlation.[9] The municipality-level clustering corrects the standard errors for within-group correlation between individuals belonging to the same labor market, and for serial correlation between monthly observations of each individual.[10] I perform the cluster adjustment of the standard errors to avoid incorrect inference by overstating the statistical significance. All placebo tests use 1 January 2014 as a placebo reform date, and estimate the expected treatment response of DI beneficiaries in the absence of a DI reform.

In addition to regular placebo tests, I also report more conservative placebo tests in section 6.3. These placebo tests use one month before and one month after the reform to ameliorate the effects of serial correlation; figure 12 and figure 13 in section 6.3 report the results from these additional placebo tests.

4.1 Data

The Norwegian Labor and Welfare Administration provided the social security register data I use in the analysis. The data contains monthly demographic and socioeconomic information on all Norwegian residents between 18 and 67 years of age, from January 2013 until December 2015. A personal ID number links the data to monthly benefit payments from the Labor and Welfare Administration to the beneficiaries. This link to the benefit payments guarantee accounting quality to the data, and it makes them perfectly suited for the analysis.

The demographic data include age, gender, marital status, country of citizenship, country of birth, number of children below 18 years of age, and the date of death. They

[9]The data contain information about the office of the Norwegian Labor and Welfare Service to which each individual belongs. For the big cities, all local offices are included in the same cluster. This ensures clustering of all individuals that live in a city that is natural thought of as a single labor market. The big city correction ensures clustering that is similar to a municipality-level clustering, since most Norwegian municipalities only have one office for the Norwegian Labor and Welfare Service.

[10]Municipality is a non-nested cluster, since individuals move between labor markets. To account for this movement, I used Stata's finite sample correction to adjust the degrees of freedom downwards by taking into account the absorbed fixed effects from the within transformation (Gormley and Matsa (2013)). The finite sample correction multiplies the estimated variance by (M/(M-1))((N-1)/(N-K)), where K is the number of estimated parameters including the fixed effects, M is the number of individuals, and N is the number of observations (Correia (2015)).

also contain welfare and labor market information, where employers report working hours on a monthly basis, start- and stop-date for the employment, and an identification number for the employer. Furthermore, the data contain information about each person's disability level, diagnosis at acceptance into the DI program, DI benefit payment information, the welfare office of each individual, and information about individuals rejected from the DI program between 2008 and 2013. Through the data, I am able to identify the treatment and the control group, as well as the intensive and extensive margin labor supply of each individual.

The dataset include the labor market information of all Norwegian citizens of working age; it therefore comprise approximately 3.1 million monthly observations and totals more than 74.8 million observations of individuals who were younger than 62 years of age in the period between January 2014 and December 2015. I restrict the sample to individuals between the age of 18 and 61 years old, to avoid any interaction with the early retirement system that allows individuals to commence early retirement at the age of 62. Individuals rejected from the DI program between 2008 and 2013, who were accepted into the DI program at appeal before January 2015, are included in the treatment group in the analysis. The extensive amount of data enables the investigation of heterogeneous labor market responses across different sub-groups, where the DD model provides precise estimates for the causal effects of the DI reform.

One problem with the data is that the implementation of the DI reform happened the same day as a reporting change of labor market information from employers to The Norwegian Labor and Welfare Administration. This reporting change affected everyone in the labor market, which means that the DI reform and the reporting change simultaneously affected the labor supply of DI beneficiaries. The control group in the analysis therefore controls for the reporting change, as well as other factors that affect the labor supply of the DI beneficiaries around the date of the reform.[11]

5 Summary statistics

Table 1 contains unconditional average statistics for the four groups in this analysis. It also contains the total number of observations from January 2014 until December 2015. Healthy individuals in column 4—those who did not receive DI benefits before January 2015 and did not apply for acceptance into the DI program between 2008 and 2013—make up the majority of all observations. Table 1 also contains average information about disability level, weekly work hours, and the percentage of individuals that are working in each of the four groups. These statistics show that rejected DI applicants, 100% DI beneficiaries, and partial DI beneficiaries work less than healthy individuals do on

[11] Appendix C provides a detailed description of the reporting change.

average. They also show that approximately 14% of all DI beneficiaries have their DI application accepted at appeal, which supports the use of rejected applicants as a control group in the analysis.

Rejected applicants are suitable as a control group for a number of reasons. Gender and age characteristics are more similar to DI beneficiaries for the rejected DI applicants than for healthy individuals. As for full and partial DI beneficiaries, women comprise the majority of the rejected DI applicant group. The rejected DI applicant group also has a higher average age, fewer working hours, and a lower wage relative to healthy individuals. More importantly, rejected DI applicants are likely to have a reduced health on average, which is the primary reason to consider them as a control group in the analysis. Overall, the summary statistics in table 1 support the choice of rejected DI applicants as a control group; further support is provided in the descriptive labor supply statistics at the end of section 5 where the parallel trends are investigated.

Table 1: Summary statistics

	(1) 100 percent DI	(2) Partial DI	(3) Rejected DI	(4) Healthy
DI level	100	58.6	5.53	.246
Work-hours if >=4hrs	19.5	20.1	28.2	34
Working	.0603	.76	.278	.73
Female	.548	.709	.553	.472
Age	49.3	52.1	46.3	38.4
Children < 18years	.314	.384	.73	.731
Rejected from DI	.139	.138	1	.00311
Yearly Income NOK	12992	191112	221552	429397
Observations	4129182	931576	229135	69579197

Sample means of all individuals younger than 62 years of age between January 2014 and December 2015.

Figure 6 and table 1 show that DI beneficiaries are primarily female and concentrated in the older age groups. However, many DI beneficiaries stay in the labor force with a reduced labor supply; table 1 indicates that work is particularly common among partial beneficiaries. In the estimation sample, there are approximately 172000 monthly observations of 100% DI beneficiaries, while the groups of partial DI beneficiaries and rejected DI applicants have approximately 38800 and 9500 monthly observations on average.

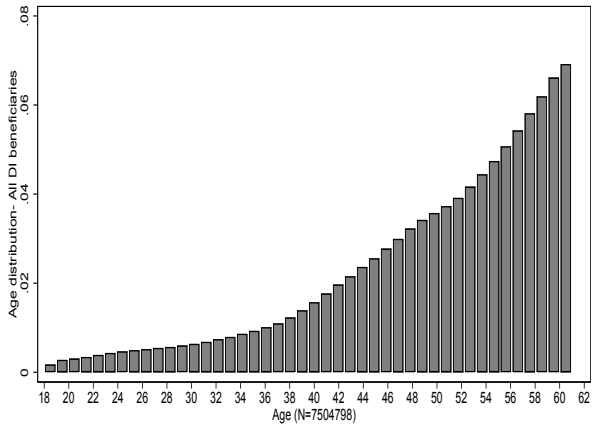

Figure 6: Age-distribution of DI beneficiary observations. N is the total number of observations in the histogram.

Figure 7 provides a summary of the average age in the four analysis groups throughout the observation period. As expected, the average age within each group does not change dramatically with time, due to strong serial correlation. Note that the average age of rejected DI applicants increases with time by definition, since this group is defined as being rejected from the DI program before January 2013.

Figure 8 shows the development in average age conditional on working more than four hours per week. The conditional average age has modest changes throughout the observation period, apart from the group of 100% DI beneficiaries; in this group, the average working age decreases after the 2015 reform. This decrease can be interpreted in one of three ways. First, the reduction could mean that young 100% DI beneficiaries are working more following the 2015 reform. Second, it could signal that fewer old 100% beneficiaries are working following the DI reform. Third, it could be the reporting change discussed in section 4.1 that affects who is registered as working after the 2015 reform. However, since only the average working age of 100% DI beneficiaries is affected in figure 8, it seems clear that this is a result of the DI reform and not the reporting change. It is therefore likely that the DI reform affected who is working in the group of 100% DI beneficiaries.

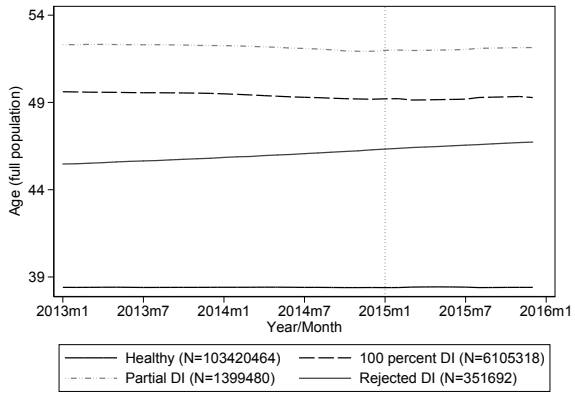

Figure 7: Average age in the treatment and control groups. N indicates the total number of observations in the group.

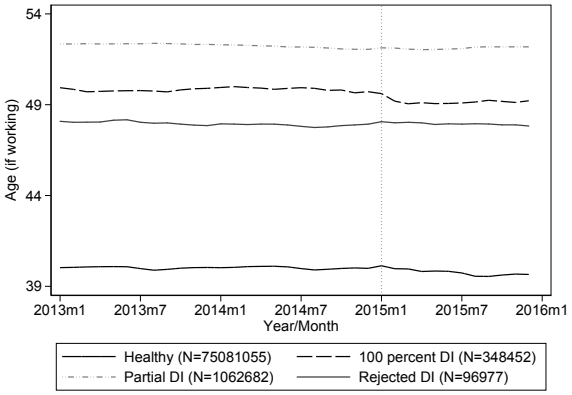

Figure 8: Average age of working invididuals in the treatment and control groups. N indicates the total number of observations in the group.

Figure 9 summarizes the development of the average disability level across time. The unconditional average disability level of all DI beneficiaries does not change with time, which indicates that the DI reform did not alter the average disability level in the treatment group. However, when the disability level is conditional on working, the average disability level increases after the implementation of the 2015 reform. This increase supports the idea that the working population in the treatment group changed after the 2015 reform. When I combine the information about the increased work incentives from the DI reform with the descriptive statistics in figures 8 and 9, it seems likely that young beneficiaries with a high disability level work more following the 2015 reform; heterogeneous labor market responses are analyzed in the empirical analysis in section 6.

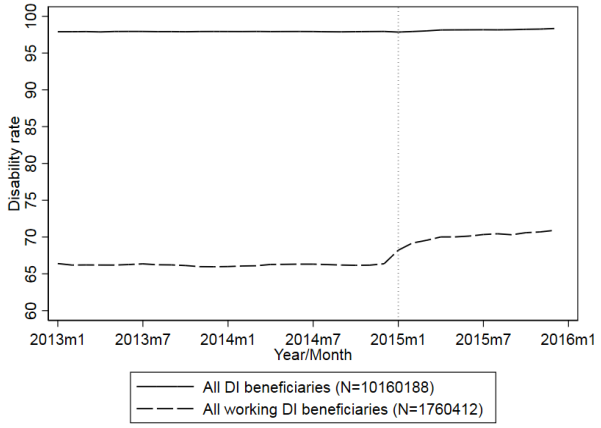

Figure 9: Average disability level for all DI beneficiaries and all working DI beneficiaries. N indicates the total number of observations in the group.

Figure 10 and 11 summarize the labor supply in the four groups throughout the observation period. These figures are important for investigating the likelihood of the identifying assumption in a DD model, which is a parallel time trend in the control and treatment group in the absence of a DI reform. Figure 10 seems to visually confirm that working hours in the control and treatment groups had a parallel trend prior to the 2015 DI reform; I further investigate the parallel trends in section 6.3. Figure 10 shows the working hours increased in all four groups around the date of the DI reform. Since the DI reform only affected beneficiaries, it seems clear that the reporting change affected the reported working hours in all four analysis groups. However, 100% DI beneficiaries clearly increased their working hours most following the DI reform, and it is difficult to visually verify any differences in the development of the working hours of partial DI beneficiaries, rejected applicants and healthy individuals following the reform.

Figure 11 shows the development in the percentage of individuals that worked more than four hours throughout the observation period. The figure indicates that the reform had a marginal total effect on the extensive margin labor supply, and it visually confirms the parallel trends assumption prior to the implementation of the DI reform. The DD analysis in section 6.2.2 estimates the extent to which there are any heterogeneous responses in the probability of working after the DI reform. It is worth noting that there is a small dip in the probability of working around the implementation of the DI reform. This dip is likely due to the reporting change discussed in section 4.1, and it seems to equally affect the individuals on the temporary DI and those individuals that are in the healthy group.

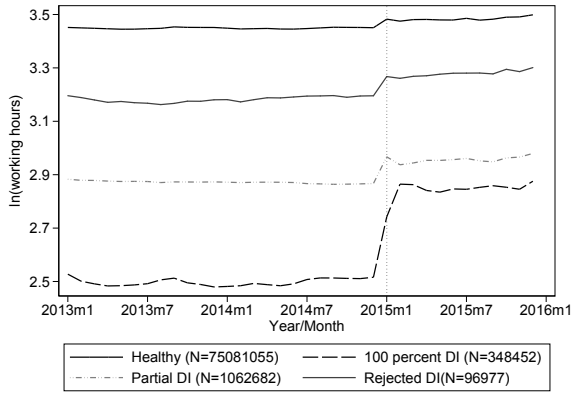

Figure 10: Development in log-transformed working hours variable. N indicates the total number of observations in the group.

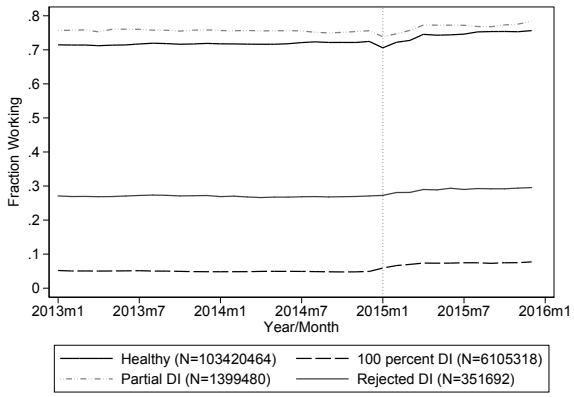

Figure 11: Development in probability of working. N indicates the total number of observations in the group.

6 Results

In this section, I analyze the labor market responses after the 2015 DI reform. I start by presenting the intensive margin labor supply analysis in section 6.1. I estimate a total effect for all DI beneficiaries across age and gender in subsection 6.1.1, while subsection 6.1.2 completes the intensive labor supply analysis by estimating the labor market responses for full and partial DI beneficiaries separately. In section 6.2 I analyze the extensive margin labor supply. Subsection 6.2.1 reports the total effect analysis for all beneficiaries, while subsection 6.2.2 presents the estimates separately for full and partial beneficiaries. I finalize the empirical analysis by presenting the placebo analysis in section 6.3.

6.1 Labor supply at the intensive margin

The intensive margin labor supply analysis estimates the percentage point change in working hours of DI beneficiaries relative to rejected DI applicants the first year following the implementation of the DI reform. I report trend-adjusted and non-trend-adjusted DD estimates throughout the analysis, and I control for individual and time fixed-effects in all regressions.

6.1.1 Total effect, intensive margin

The estimates for the total intensive labor market response after the 2015 DI reform are reported in table 2. These regressions pool the observations of partial and full DI beneficiaries, to estimate the overall effects of the DI reform. Columns 1-3 report the non-trend-adjusted DD estimates, while columns 4-6 contain the estimates from the trend-adjusted model. The first panel reports the estimated percentage point change in working hours for all DI beneficiaries between ages 18 and 61, while the last three panels separate the total estimates into smaller age-categories, to investigate potential heterogeneous responses across groups categorized by gender and age.

The estimated total labor supply response after the DI reform is a 2.5 percentage point increase in working hours of all DI beneficiaries in the trend-adjusted model. This effect splits into a 4.5 percentage point increase for men, and a 1.4 percentage point increase for women. Without a trend adjustment, the estimated total effect is a 0.4 percent point increase in the working hours, comprising a 3.4 percentage point increase for men and a 1.1 percentage point decrease for women. The last three panels in table 2 indicate that young DI beneficiaries have the highest point estimate for the intensive labor market response. The results also indicate that men have consistently higher estimated effects relative to women in both models, where the estimated effects are statistically significant for men. The estimated effects are similar across both specifications of the DD model.

In all, there is a clear indication that the economic incentives program in the 2015 DI reform had a positive effect on the working hours of male DI beneficiaries. There is also a positive and statistically significant overall change in the working hours of all beneficiaries, where the effects decrease with age.

Table 2: 2014/2015 Difference-in-difference estimates: Percentage point change in working hours. Dependent variable: log (work hours). Control group: Rejected DI applicants.

	All DI vs. Rejected					
	(1) Female	(2) Male	(3) Pooled	(4) Female	(5) Male	(6) Pooled
Age 18-61						
Diff-in-diff estimate	-0.011	0.034**	0.004	0.014*	0.045***	0.025***
	(0.0084)	(0.0144)	(0.0070)	(0.0078)	(0.0125)	(0.0066)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[690875]	[330081]	[1020956]	[690875]	[330081]	[1020956]
Clusters	435	435	437	435	435	437
Age 18-29						
Diff-in-diff estimate	-0.048	0.135***	0.047	-0.037	0.104*	0.039
	(0.0677)	(0.0427)	(0.0404)	(0.0771)	(0.0601)	(0.0400)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[8817]	[10582]	[19399]	[8817]	[10582]	[19399]
Clusters	217	243	285	217	243	285
Age 30-49						
Diff-in-diff estimate	-0.007	0.062***	0.016	0.017	0.070***	0.036***
	(0.0139)	(0.0164)	(0.0115)	(0.0116)	(0.0151)	(0.0094)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[230099]	[112448]	[342547]	[230099]	[112448]	[342547]
Clusters	429	420	434	429	420	434
Age 50-61						
Diff-in-diff estimate	-0.004	-0.005	-0.004	0.015	0.016	0.016
	(0.0126)	(0.0241)	(0.0136)	(0.0116)	(0.0186)	(0.0096)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[451959]	[207051]	[659010]	[451959]	[207051]	[659010]
Clusters	434	429	436	434	429	436

Intensive margin DD estimates using individual and monthly fixed effects, with an observation window of +/- 12 months relative 1 January 2015. The first panel contains pooled regressions for all individuals 18-61 years of age, where each regression is run separately for women, men, and both genders. The next three panels contain a subgroup analysis focusing on gender and age effects. The outcome variable is a log-transformed working hours variable, and (standard errors) are clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01

6.1.2 Intensive margin heterogeneity analysis by the level of disability

Columns 1-6 in table 3 report the estimated percentage point change in working hours for 100% DI beneficiaries after the 2015 reform. The estimates indicate that full DI beneficiaries increased their working hours by 8.2 percentage points, comprising an estimated 10.4 percentage point increase for men and a 6.9 percentage point increase for women. The model without the trend adjustment estimates similar effects, and it seems clear that the DI reform had a pronounced overall effect on the working hours of 100% DI beneficiaries. Full DI beneficiaries between ages 18–29 had an estimated 5.3 percentage point increase in the working hours after the DI reform; however, this estimate is not statistically significant at a conventional level. With the lacking power in the youngest age category, it is also likely that the reform did not affect the average working hours of this group. For the 30-49 and the 50-61 age groups, the estimates show that the working hours increased by approximately 8 percentage points in both groups, which is similar to the estimated total effect in the top panel. The analysis suggests that there is no clear age effect for 100% DI beneficiaries, but there is a gender effect where the point estimate for the labor market response is consistently higher for men than for women. Overall, the analysis indicates a pronounced increase in the working hours of 100% DI beneficiaries after the 2015 DI reform.

Columns 7-12 in table 3 report the estimated intensive margin labor market response of partial DI beneficiaries. They saw an estimated 1.2 percentage point increase in working hours after the DI reform, comprising a 1.6 percentage point increase for men and a 0.6 percentage point increase for women. For partial DI beneficiaries, the statistically significant effects are concentrated in the 30-49 age-group; this group increased their working hours by 2.2 percentage points. Overall, partial beneficiaries did not change their working hours to a large extent after the DI reform.

The heterogeneity analysis by the level of disability in table 3 indicates that 100% DI beneficiaries had the largest intensive labor market response after the 2015 DI reform. Partial beneficiaries saw a modest increase in working hours in the 30-49 age-group, and I can think of three mechanisms for why the labor market response is lower among partial beneficiaries. First, the old DI system might have had a more restrictive SGA-threshold for full beneficiaries than it did for partial beneficiaries. Second, the lower effects for partial beneficiaries could signal that this group had a low unused potential labor supply prior to the implementation of the 2015 reform; this would make economic incentives an inefficient tool for a labor supply increase in this group. Third, it could be that the partial beneficiaries avoids revealing their unused potential labor supply, because they want to apply for full DI benefits; a high work capacity would make them ineligible for 100% DI benefits.

A key result in the intensive margin labor supply analysis, is that men respond more strongly than women in both DI beneficiary groups. This could be because it is easier to mobilize men with economic incentives, even if poor health characterize all the individuals in both DI groups. The larger effects for men at the intensive margin are similar to prior estimated behavior at the extensive margin in studies such as Kostøl and Mogstad (2014). It is unclear why men have the highest labor market response, but one possible explanation might be demand side effects: it might be easier to vary the work hours in male dominated industries. There is also a clear age effect at the intensive margin overall, where partial DI beneficiaries are the drivers behind this decreasing overall age effect. The lack of an age effect for 100% DI beneficiaries contrasts with prior estimated extensive labor market responses, which decreases with age for full and partial DI beneficiaries (Kostøl and Mogstad (2014); Koning and van Sonsbeek (2017)). In total, the analysis indicates that the 2015 DI reform significantly increased the working hours of 100% DI beneficiaries.

Table 3: 2014/2015 Difference-in-difference estimates: Percentage point change in working hours. Dependent variable: log (work hours). Control group: Rejected DI applicants.

	100 percent DI vs. Rejected						Partial DI vs. Rejected					
	(1) Female	(2) Male	(3) Pooled	(4) Female	(5) Male	(6) Pooled	(7) Female	(8) Male	(9) Pooled	(10) Female	(11) Male	(12) Pooled
Age 18-61												
Diff-in-diff estimate	0.065***	0.118***	0.086***	0.069***	0.104***	0.082***	-0.023***	-0.006	-0.015**	0.006	0.016	0.012*
	(0.0114)	(0.0203)	(0.0112)	(0.0112)	(0.0165)	(0.0096)	(0.0085)	(0.0128)	(0.0068)	(0.0079)	(0.0119)	(0.0066)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[172816]	[140033]	[312849]	[172816]	[140033]	[312849]	[558876]	[213001]	[771877]	[558876]	[213001]	[771877]
Clusters	432	430	435	432	430	435	433	425	435	433	425	435
Age 18-29												
Diff-in-diff estimate	-0.030	0.149***	0.064	-0.004	0.102	0.053	-0.070	0.088	0.012	-0.084	0.109	0.009
	(0.0701)	(0.0454)	(0.0420)	(0.0805)	(0.0625)	(0.0430)	(0.0732)	(0.0654)	(0.0489)	(0.0786)	(0.0776)	(0.0452)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[5778]	[8479]	[14257]	[5778]	[8479]	[14257]	[4288]	[3532]	[7820]	[4288]	[3532]	[7820]
Clusters	191	224	273	191	224	273	144	133	198	144	133	198
Age 30-49												
Diff-in-diff estimate	0.060***	0.126***	0.083***	0.059***	0.118***	0.080***	-0.021	0.024	-0.004	0.008	0.040**	0.022**
	(0.0181)	(0.0220)	(0.0143)	(0.0164)	(0.0191)	(0.0141)	(0.0142)	(0.0177)	(0.0121)	(0.0119)	(0.0163)	(0.0093)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[70079]	[54635]	[124714]	[70079]	[54635]	[124714]	[179762]	[68507]	[248269]	[179762]	[68507]	[248269]
Clusters	411	392	426	411	392	426	421	383	427	421	383	427
Age 50-61												
Diff-in-diff estimate	0.079***	0.094***	0.087***	0.080***	0.086***	0.083***	-0.013	-0.042**	-0.022*	0.008	-0.012	0.003
	(0.0156)	(0.0307)	(0.0182)	(0.0154)	(0.0235)	(0.0132)	(0.0126)	(0.0213)	(0.0127)	(0.0117)	(0.0173)	(0.0094)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[96959]	[76919]	[173878]	[96959]	[76919]	[173878]	[374826]	[140962]	[515788]	[374826]	[140962]	[515788]
Clusters	421	419	432	421	419	432	431	410	433	431	410	433

Intensive margin DD estimates using individual and monthly fixed effects, with an observation window of +/- 12 months relative 1 January 2015. The first panel contains pooled regressions for all individuals aged 18-61, where each regression is run separately for women, men and both genders. The next three panels contain a subgroup analysis focusing on gender and age effects. The outcome variable is a log-transformed working hours variable, and (standard errors) are clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01

6.2 Labor supply at the extensive margin

The extensive margin labor supply analysis estimates the percentage point change in the probability that DI beneficiaries worked more than four hours, relative to rejected DI applicants the first year after the 2015 DI reform. I report both trend-adjusted and non-trend-adjusted DD estimates, and all regressions control for individual and time fixed effects.

6.2.1 Total effect, extensive margin

Columns 1-6 in table 4 report the estimates from regressions that pool the observations of partial and full DI beneficiaries, to estimate a total extensive margin labor market response after the DI reform. Columns 1-3 contain the estimates from the DD model without a trend-adjustment, while columns 4-6 report the estimates from the trend-adjusted model. The first panel reports the total estimates for all DI beneficiaries, and the three panels at the bottom of the table group the estimated effects into smaller age categories.

The estimated total effect of the DI reform is no change in the probability of working for all DI beneficiaries. The total effect is only statistically significant for women, who decreased their extensive margin labor supply by 0.7 percentage points. However, men increased their extensive labor supply by 0.6 percentage points, and the total estimate indicates that the DI reform did not affect the overall extensive margin labor supply of the DI beneficiaries.

The estimated total effect decreases with age in both DD models. The estimates indicate a 2.2 percentage point increase in the probability of working in the 18-29 age group, where men drives this change with a 3 percentage point increase in the probability of working. In the oldest age group, the extensive margin labor supply reduced by 0.9 percentage points. Overall it seems clear that the total effect of the 2015 DI reform on the extensive labor supply is small and not statistically significant, and the model without the trend-adjustment finds a small negative overall response to the DI reform.

Table 4: 2014/2015 Difference-in-difference estimates: Percentage point change in labor market participation Control group: Rejected DI applicants

	All DI vs. Rejected					
	(1) Female	(2) Male	(3) Pooled	(4) Female	(5) Male	(6) Pooled
Age 18-61						
Diff-in-diff estimate	-0.016***	-0.006*	-0.012***	-0.007**	0.006*	-0.001
	(0.0024)	(0.0034)	(0.0021)	(0.0027)	(0.0031)	(0.0022)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[3013871]	[2199877]	[5213748]	[3013871]	[2199877]	[5213748]
Clusters	436	440	440	436	440	440
Age 18-29						
Diff-in-diff estimate	0.011	-0.004	0.003	0.013	0.030***	0.022***
	(0.0096)	(0.0118)	(0.0070)	(0.0093)	(0.0104)	(0.0067)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[123333]	[158272]	[281605]	[123333]	[158272]	[281605]
Clusters	404	413	423	404	413	423
Age 30-49						
Diff-in-diff estimate	-0.011***	-0.006	-0.009***	0.001	0.006	0.003
	(0.0040)	(0.0050)	(0.0031)	(0.0046)	(0.0046)	(0.0034)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[1045362]	[803633]	[1848995]	[1045362]	[803633]	[1848995]
Clusters	435	435	438	435	435	438
Age 50-61						
Diff-in-diff estimate	-0.018***	-0.001	-0.011***	-0.017***	0.002	-0.009***
	(0.0036)	(0.0051)	(0.0033)	(0.0036)	(0.0040)	(0.0026)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[1845176]	[1237972]	[3083148]	[1845176]	[1237972]	[3083148]
Clusters	436	439	439	436	439	439

Extensive margin DD estimates using individual and monthly fixed effects, with an observation window of +/- 12 months relative 1 January 2015. The first panel contains pooled regressions for all individuals 18-61 years of age, where each regression is run separately for women, men and both genders. The next three panels contain a subgroup analysis focusing on gender and age effects. The outcome variable is a dummy variable for working more than four hours, and (standard errors) are clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01

6.2.2 Extensive margin heterogeneity analysis by the level of disability

In this section, I separate the extensive labor market response for full and partial DI beneficiaries. Columns 1-6 of table 5 report the estimated percentage point change in the probability of working more than four hours for 100% DI beneficiaries after the 2015 DI reform. The estimated effects in columns 4-6 indicate the DI reform did not cause a change in the labor force participation of 100% DI beneficiaries. The 0.9 percentage point increase in the labor market participation of 100% DI beneficiary men, is offset by a 0.5 percentage point decrease among the females in the same group. The estimated effects decrease with age for both genders in trend-adjusted model, and column 5 indicates that men in the 18-29 age group had a 3.1 percentage point increase in the probability of working. In total, the DI reform seems to not have affected the overall labor force participation of 100% DI beneficiaries.

Column 12 in table 5 shows a 1.2 percentage point decrease in the probability that partial beneficiaries work after the DI reform. This overall decrease is driven by the 1.9 percentage point decrease in the labor force participation in the 50-61 age-group. The extensive margin estimates indicate that the DI reform slightly reduced the probability that partial DI beneficiaries work after the reform.

Overall, the heterogeneity analysis shows that the extensive labor market response after the DI reform has been modest for both 100% and partial DI beneficiaries. For 100 percent DI beneficiaries, men increased their extensive margin labor supply, and the largest response occurred in the youngest age category. The estimated effects also decrease with age for both males and females claiming full DI benefits. Among female 100% beneficiaries we see a statistically significant reduction in the labor force participation of 1.5 percentage points in the oldest age group, but this effect is offset by the increase in the labor force participation of young men. Partial DI beneficiaries slightly reduced their extensive margin labor supply after the DI reform , and the overall effect is a statistically significant 1.2 percentage point reduction in the labor force participation. The estimates therefore indicate that the extensive labor market response to the changes in the economic incentive structure is strongest among young men that claim full DI benefits, while partial DI beneficiaries slightly reduced their extensive margin labor supply.

Table 5: 2014/2015 Difference-in-difference estimates: Percentage point change in labor market participation. Control group: Rejected DI applicants

	100 percent DI vs. Rejected						Partial DI vs. Rejected					
	(1) Female	(2) Male	(3) Pooled	(4) Female	(5) Male	(6) Pooled	(7) Female	(8) Male	(9) Pooled	(10) Female	(11) Male	(12) Pooled
Age 18-61												
Diff-in-diff estimate	-0.008***	-0.000	-0.005**	-0.005*	0.009***	0.001	-0.029***	-0.029***	-0.029***	-0.012***	-0.018***	-0.012***
	(0.0025)	(0.0035)	(0.0021)	(0.0027)	(0.0031)	(0.0023)	(0.0028)	(0.0041)	(0.0024)	(0.0032)	(0.0038)	(0.0025)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[2357891]	[1931318]	[4289209]	[2357891]	[1931318]	[4289209]	[782141]	[370443]	[1152584]	[782141]	[370443]	[1152584]
Clusters	436	439	439	436	439	439	434	436	437	434	436	437
Age 18-29												
Diff-in-diff estimate	0.011	-0.004	0.003	0.013	0.031***	0.022***	0.008	-0.002	0.003	-0.000	-0.011	-0.002
	(0.0095)	(0.0117)	(0.0069)	(0.0094)	(0.0104)	(0.0067)	(0.0247)	(0.0346)	(0.0198)	(0.0229)	(0.0332)	(0.0189)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[118561]	[155314]	[273875]	[118561]	[155314]	[273875]	[12992]	[11687]	[24679]	[12992]	[11687]	[24679]
Clusters	402	412	423	402	412	423	234	226	296	234	226	296
Age 30-49												
Diff-in-diff estimate	-0.007*	-0.002	-0.005	0.002	0.010**	0.006	-0.019***	-0.034***	-0.023***	-0.002	-0.019***	-0.006
	(0.0041)	(0.0051)	(0.0032)	(0.0044)	(0.0047)	(0.0035)	(0.0048)	(0.0067)	(0.0038)	(0.0060)	(0.0064)	(0.0044)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[839858]	[722098]	[1561956]	[839858]	[722098]	[1561956]	[268649]	[131982]	[400631]	[268649]	[131982]	[400631]
Clusters	435	434	437	435	434	437	426	413	430	426	413	430
Age 50-61												
Diff-in-diff estimate	-0.009**	0.005	-0.003	-0.015***	0.006	-0.006**	-0.031***	-0.020***	-0.026***	-0.023***	-0.019***	-0.019***
	(0.0036)	(0.0051)	(0.0032)	(0.0036)	(0.0040)	(0.0026)	(0.0038)	(0.0053)	(0.0034)	(0.0041)	(0.0042)	(0.0029)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[1399472]	[1053906]	[2453378]	[1399472]	[1053906]	[2453378]	[500500]	[226774]	[727274]	[500500]	[226774]	[727274]
Clusters	436	439	439	436	439	439	433	429	435	433	429	435

Extensive margin DD estimates using individual and monthly fixed effects, with an observation window of +/- 12 months relative 1 January 2015. The first panel contains pooled regressions for all individuals 18-61 years of age, where each regression is run separately for women, men and both genders. The next three panels contain a subgroup analysis focusing on gender and age effects. The outcome variable is a dummy variable for working more than four hours, and (standard errors) are clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01

6.3 Placebo tests

Figure 12 and 13 present the beta estimates from the following DD model:

$$Y_{it} = \alpha_i + \alpha_1 post_t + \alpha_2 treat_i + \beta(treat_i * post_t) + \epsilon_{it}$$

January 2014 is a permanent pre-reform period in the DD model, and I use each of the subsequent months as post-periods. Similarly to the regressions in section 6, the DD model compares the labor supply of all DI beneficiaries with rejected DI applicants, and the model clusters observations at the municipality level. The Feb14 DD estimate in figure 12 uses January 2014 as a pre-period and February 2014 as the post-period, while Jan15 uses January 2014 as the pre-period and January 2015 as the post-period. Figure 12 and figure 13 are used to inspect whether serial correlation drives the estimated effects in section 6. The figures additionally test for whether the treatment and the control group have a parallel trend prior to the 2015 DI reform. The DD model should find no statistically significant estimates prior to Jan15 if the treatment and the control group have a parallel trend prior to the reform.

Figure 12 seems to indicate that serial correlation does not drive the DD estimates in the intensive margin labor supply analysis, since the estimates in the figure only use one pre-period and one post-period. The figure also finds no statistically significant DD estimates prior to the implementation of the DI reform, but there is a slight indication of a downward trend in the estimated effects prior to the DI reform. This could indicate the need for a trend adjustment in the panel data analysis analysis of the change in working hours, to remove a potential source of bias.

Figure 13 similarly tests for whether serial correlation drives the estimated effects in the extensive margin labor supply analysis in section 6.2. The extensive margin placebo tests in figure 13 find that the statistically significant effects hit after the implementation of the DI reform, and the DD estimates trend slightly downward in figure 13 prior to the implementation of the DI reform.

In sum, the placebo tests in figure 12 and 13 provide support for the causal analysis in section 6. The figures indicate that a trend-adjusted model might be necessary in the panel data analysis, due to the slight downward trend in the DD estimates in both figures prior to the implementation of the reform. A trend adjustment will loosen the parallel trend assumption and remove a potential source of bias, even if the causal estimates in section 6 are not substantially affected by this trend-adjustment.

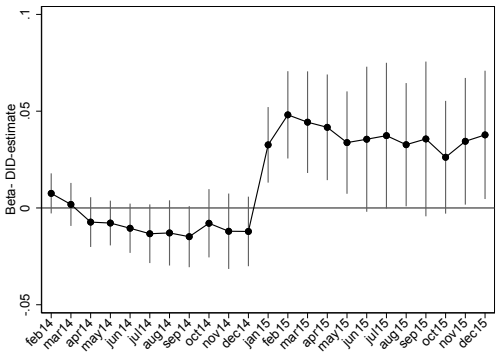

Figure 12: Difference-in-difference placebo tests, intensive margin labor supply.

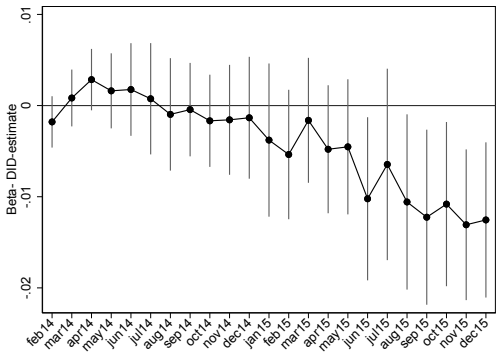

Figure 13: Difference-in-difference placebo tests, extensive margin labor supply.

6.3.1 Intensive margin placebo tests

In appendix A, I report the placebo estimates for table 2, 3, 4, and 5. All placebo estimates use 1 January 2014 as a placebo reform date and rejected DI applicants as the control group.

Columns 1-6 in table 6 report the placebo estimates for the total intensive labor supply analysis in table 2. The placebo regressions without a trend adjustment in columns 1-3 of table 6 capture some statistically significant effects. However, when a trend adjustment is included in columns 4-6, the statistically significant placebo effects disappear. Since the placebo estimates from the model without a trend-adjustment in table 6 are negative and statistically significant, it is likely that this model produce a bias in the reform estimates. Figure 12 and 13 therefore correctly indicate that a trend adjustment is needed to correct for a potential source of bias; I report the estimates without a trend adjustment in table 2 for transparency. The main model in the analysis is therefore the richest DD model, which loosens the parallel trend assumption.

Table 7 in appendix A presents the intensive margin placebo estimates for the heterogeneity analysis in section 6.1.2. The model without a trend adjustment again finds statistically significant placebo estimates for both full and partial DI beneficiaries, but these significant effects disappear when a trend adjustment is included in the DD model. The trend adjusted model therefore seems to correct for the minor difference in trend prior to the implementation of the DI reform. In total, the trend-adjusted estimates in table 6 and 7 support the likelihood that the trend-adjusted estimates in table 2 and 3 are the causal intensive margin labor market responses following the 2015 DI reform.

6.3.2 Extensive margin placebo tests

Columns 1-6 in table 8 report the placebo estimates for the total extensive labor market analysis in table 4. The non-trend-adjusted placebo tests capture statistically significant effects, which predominantly disappear when a trend adjustment is included in the DD model. There is a small 0.4 percentage point overall effect in the top panel that is statistically significant at at a 5% level. This point estimate is at the same level as the placebo tests that are not statistically significant; however, the estimated variance is extremely low due to the 5.1 million observations in the regression. Overall, it seems like the trend adjustment removes a source of bias in the extensive labor market analysis, and the estimates without a trend adjustment in table 4 are reported for transparency. The main DD model is therefore loosening the parallel trends assumption by correcting for a difference in trend.

The trend-adjusted placebo tests for the extensive labor market responses of full DI beneficiaries are reported in columns 4-6 in table 9. The test finds a small but significant overall effect of 0.5 percentage points; this significant effect is at the same level as the placebo tests, but the standard errors are small due to the 4.2 million observations in the regression. In columns 10-12, partial DI beneficiaries have no statistically significant placebo estimates, and the trend-adjusted placebo estimates in table 9 seem to support the likelihood that the trend-adjusted estimates in table 5 are the causal extensive margin labor market responses following the DI reform.

In total, the placebo estimates in tables 6, 7, 8, and 9 support the use of a trend-adjusted model in this analysis. Figure 12 and 13 also indicate that a linear trend might capture the minor difference in the pre-reform trend. As is usual in most empirical analysis, the main DD model in this analysis is the one with the richest specification, and the estimates from the trend-adjusted model are therefore preferred throughout this analysis.

7 Conclusion

In this study, I have completed an intensive and extensive margin labor market analysis after a 2015 DI reform on the full population in Norway. Much of the prior literature on the economic incentive approach to DI reform has focused on extensive margin labor market responses, while I use monthly data on reported working hours to estimate both the intensive and the extensive margin labor market responses following the reform. The DI reform implemented an economic incentive program to increase the labor supply of DI beneficiaries; the responses along both labor supply margins are important for the long-term viability of the DI program.

I estimate the ATE of the DI reform in a DD framework, because the ATE is an important parameter for a policy maker; this study adds to a literature that has focused primarily on LATE estimates in a RD or an instrumental variable framework. The results from the analysis indicate that the ATE at the extensive margin is significantly lower than prior estimated LATE estimates in studies such as that of Kostøl and Mogstad (2014); they find a 9 percentage point increase in the extensive margin labor supply, which is measured as the probability of earning an income above the SGA threshold. The ATE should be lower than a LATE estimate by default, but the average extensive margin labor market responses in this analysis are much smaller than prior estimated local effects. My analysis therefore supports Autor and Duggan (2006), who assess that it is a challenge to activate individuals who have already been awarded a permanent DI benefit. The problem of activating DI beneficiaries might be an issue of demand for the labor provided by DI beneficiaries. It might be difficult for DI beneficiaries to enter the labor market in periods when their health permits a higher labor supply, because employers may be hesitant to hire individuals that enter and exit the labor market with variations in health.

Because the 2015 reform did not alter the DI benefit level, it is natural to think that the reform should not greatly affect the extensive margin labor supply of DI beneficiaries, and this analysis confirms this assessment. My extensive margin labor market analysis confirms the age-effects found in Kostøl and Mogstad (2014) and Koning and van Sonsbeek (2017), where young DI beneficiaries respond most strongly to economic incentives. I also find that economic incentives greatly affect the average working hours of DI beneficiaries, which is an important result in this analysis. The total estimated effects on working hours are positive and decreasing with age, where partial beneficiaries drives this age effect. The larger intensive margin labor market response might occur because it is easier for individuals who are already in the labor market to increase their working hours with an existing employer, relative to the challenge of finding a new employer for someone on the outside of the labor market. This finding seems to be confirmed by figure 12 and 13, where the labor market response is faster in figure 12 than what it is in figure 13.

The changes in working hours are large and positive in most sub-groups following the 2015 DI reform. 100% DI beneficiaries increased their working hours by as much as 8.2 percentage points, on average. The removal of the discontinuities in the budget constraint therefore greatly affected individuals that at some point had been evaluated to have fully lost their ability to gain an income. This supports the argument Moore (2015) makes, where the impermanence of some disability conditions makes it essential to structure the DI system with a goal of maximizing labor supply of beneficiaries over time. This allows individuals to change their labor supply when their health situation changes. The analysis favors a gradual reduction of DI benefits against an increasing

income, relative to a system that has sharp discontinuities in the budget constraint. The removal of the discontinuities seems to reduce the ratchet effects in the DI system; it allows DI beneficiaries to change their working hours according to the variations in their health.

Another important result of the analysis is that men respond most strongly to the changes in the incentive structure at the intensive margin. This is interesting, since early labor market studies found that females have a more elastic labor supply than men; see e.g., Heckman (1993) and Evers et al. (2008). However, most DI beneficiaries have a low labor supply, just as females did in the early labor market studies. It therefore seems like individuals with a low number of working hours have a more elastic labor supply, and that gender is not necessarily tied to the lower elasticity. The estimated labor supply increases in working hours of DI beneficiaries is consistently higher for men across all age groups. However, the mechanisms behind this gender effect are not clear. One explanation could be that men have a higher salary prior to becoming disabled, which causes them to have a lower real replacement rate in the DI program. This lower real replacement rate gives men a lower average benefit reduction with an increased salary after the DI reform, which leaves them with a higher hourly wage relative to females.

A very rough calculation indicates the total increased labor supply of the DI beneficiaries following the 2015 DI reform is approximately equal to a labor supply increase of 646 full-time positions among the 311,145 DI beneficiaries.[12] This builds on the assumption that the 0.001 estimate in the extensive labor supply analysis is equal to null. The whole labor supply increase therefore comes from the intensive margin labor market response, where I only assume that the causal effect of the DI reform is correctly estimated. The value gained from the intensive margin labor market response emphasizes the importance of encouraging an increased labor supply along both the intensive and extensive margin; the DI beneficiaries will pay an income tax for the salary gained from the increased working hours, and their DI benefits will be reduced because of the reduction model in the new system.

The analysis indicates that continuously offering DI beneficiaries economic incentives to utilize their potential labor supply is particularly effective at the intensive labor supply. The analysis also finds that economic incentives are less efficient at encouraging DI beneficiaries to enter the labor force, on average. However, young men with a high grade of disability respond strongly at both the intensive and the extensive margin, which indicates that poor health is not necessarily a permanent condition. Prior health might also

[12]Value of intensive margin labor market response: (Average working hours all working DI beneficiaries December 2014 * Pecentage point increase in the working hours after 2015 reform * Number of working DI beneficiaries December 2014) / Number of hours in a full work week) = ((18.39 * 0.025 * 52675)/ 37.5) = 645.8)

be a poor predictor of the unused potential labor supply of the DI beneficiaries, which emphasizes the importance of creating a DI system that continuously encourage all DI beneficiaries to maximize their labor supply. This can reduce the long-term costs and increase the long-term viability of the DI program.

References

Aakvik, A., Bjorvatn, A., Holmås, T. H., and Steckmest, E. (2001). *Endringer i regler for uføre- og alderspensjonister : hvordan påvirkes arbeidstilbudet for disse gruppene? (Changes in the rules of disability and retirement beneficiaries: how is the labor supply of these groups affected?)*, volume nr 19/01 of *SNF-rapport (trykt utg.)*. Stiftelsen for samfunns- og næringslivsforskning, Bergen.

Autor, D. and Duggan, M. (2003). The rise in the disability rolls and the decline in unemployment. *The 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.

Benıtez-Silva, H., Buchinsky, M., and Rust, J. (2006). Induced entry effects of a $1 for $2 offset in SSDI benefits. *Manuscript, SUNY-Stony Brook, UCLA, and University of Maryland*.

Blundell, R., Bozio, A., and Laroque, G. (2013). Extensive and intensive margins of labour supply: Work and working hours in the US, the UK and France. *Fiscal Studies*, 34(1):1–29.

Blyth, B. (2006). Incapacity benefit reforms: Pathways to Work pilots performance and analysis. *DWP Working Paper*, 26.

Borghans, L., Gielen, A. C., and Luttmer, E. F. (2014). Social support substitution and the earnings rebound: Evidence from a regression discontinuity in disability insurance reform. *American Economic Journal: Economic Policy*, 6(4):34–70.

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

Bound, J., Cullen, J. B., Nichols, A., and Schmidt, L. (2004). The welfare implications of increasing disability insurance benefit generosity. *Journal of Public Economics*, 88(12):2487–2514.

Bound, J., Lindner, S., and Waidmann, T. (2010). Reconciling findings on the employment effect of disability insurance. Technical report, University of Michigan, Michigan Retirement Research Center.

Campolieti, M. (2004). Disability insurance benefits and labor supply: Some additional evidence. *Journal of Labor Economics*, 22(4):863–889.

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(3-4):306–316.

Correia, S. (2015). Singletons, cluster-robust standard errors and fixed effects: A bad mix. *Technical Note, Duke University*.

Dahl, G. B., Kostøl, A. R., and Mogstad, M. (2014). Family welfare cultures. *The Quarterly Journal of Economics*, 129(4):1711–1752.

Duggan, M. and Imberman, S. A. (2009). Why are the disability rolls skyrocketing? The contribution of population characteristics, economic conditions, and program generosity. In *Health at older ages: The causes and consequences of declining disability among the elderly*, pages 337–379. University of Chicago Press.

Evers, M., De Mooij, R., and Van Vuuren, D. (2008). The wage elasticity of labour supply: a synthesis of empirical estimates. *De Economist*, 156(1):25–43.

French, E. and Song, J. (2014). The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy*, 6(2):291–337.

Gormley, T. A. and Matsa, D. A. (2013). Common errors: How to (and not to) control for unobserved heterogeneity. *The Review of Financial Studies*, 27(2):617–661.

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

Haveman, R. H. and Wolfe, B. L. (1984). The decline in male labor force participation: comment. *Journal of Political Economy*, 92(3):532–541.

Heckman, J. J. (1993). What has been learned about labor supply in the past twenty years? *The American Economic Review*, 83(2):116–121.

Hill, M. J., Maestas, N., and Mullen, K. J. (2016). Employer accommodation and labor supply of disabled workers. *Labour economics*, 41:291–303.

Koning, P. and van Sonsbeek, J.-M. (2017). Making disability work? The effects of financial incentives on partially disabled workers. *Labour Economics*, 47:202–215.

Kostøl, A. and Mogstad, M. (2014). How financial incentives induce disability insurance recipients to return to work. *American Econonic Review*, 104(2):624–655.

Low, H. and Pistaferri, L. (2015). Disability insurance and the dynamics of the incentive insurance trade-off. *American Economic Review*, 105(10):2986–3029.

Maestas, N., Mullen, K. J., and Strand, A. (2013). Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt. *American Economic Review*, 103(5):1797–1829.

Moffitt, R. A. (1999). Explaining welfare reform: Public choice and the labor market. *International Tax and Public Finance*, 6(3):289–315.

Moore, T. J. (2015). The employment effects of terminating disability benefits. *Journal of Public Economics*, 124:30–43.

OECD (2009). Pathways onto (and off) disability benefits: Assessing the role of policy and individual circumstances. *Tackling the Jobs Crisis. OECD Employment Outlook*.

OECD (2018). Public spending on incapacity - total, % of gdp, 2015 or latest available. https://data.oecd.org/socialexp/public-spending-on-incapacity.htm [Online; Accessed on 26 March 2018].

Parsons, D. O. (1980). The decline in male labor force participation. *Journal of Political Economy*, 88(1):117–134.

Wachter, T. v., Song, J., and Manchester, J. (2011). Trends in employment and earnings of allowed and rejected applicants to the social security disability insurance program. *American Economic Review*, 101(7):3308–3329.

Weathers, R. R. and Hemmeter, J. (2011). The impact of changing financial work incentives on the earnings of social security disability insurance (SSDI) beneficiaries. *Journal of Policy Analysis and Management*, 30(4):708–728.

Appendix A

Table 6: 2013/2014 Difference-in-difference placebo estimates: Percentage point change in working hours. Dependent variable: Log (work hours). Control group: Rejected DI applicants.

	All DI vs. Rejected					
	(1) Female	(2) Male	(3) Pooled	(4) Female	(5) Male	(6) Pooled
Age 18-61						
Diff-in-diff estimate	-0.021***	-0.020***	-0.020***	-0.001	0.002	-0.000
	(0.0059)	(0.0062)	(0.0045)	(0.0091)	(0.0064)	(0.0068)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[675395]	[312972]	[988367]	[675395]	[312972]	[988367]
Clusters	435	434	437	435	434	437
Age 18-29						
Diff-in-diff estimate	0.038	-0.069**	-0.018	0.014	0.007	0.010
	(0.0290)	(0.0293)	(0.0217)	(0.0461)	(0.0330)	(0.0307)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[7430]	[8939]	[16369]	[7430]	[8939]	[16369]
Clusters	190	218	266	190	218	266
Age 30-49						
Diff-in-diff estimate	-0.036***	-0.027***	-0.033***	-0.008	-0.005	-0.007
	(0.0080)	(0.0095)	(0.0060)	(0.0130)	(0.0092)	(0.0088)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[216927]	[103599]	[320526]	[216927]	[103599]	[320526]
Clusters	427	413	432	427	413	432
Age 50-61						
Diff-in-diff estimate	-0.007	-0.002	-0.005	0.004	0.008	0.005
	(0.0071)	(0.0086)	(0.0057)	(0.0072)	(0.0089)	(0.0064)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[451038]	[200434]	[651472]	[451038]	[200434]	[651472]
Clusters	434	426	436	434	426	436

Intensive margin DD estimates using individual and monthly fixed effects, with an observation window of +/- 12 months relative 1 January 2014. The first panel contains pooled regressions for all individuals 18-61 years of age, where each regression is run separately for women, men and both genders. The next three panels contain a subgroup analysis focusing on gender and age effects. The outcome variable is a log-transformed working hours variable, and (standard errors) are clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01

Table 7: 2013/2014 Difference-in-difference placebo estimates: Percentage point change in working hours. Dependent variable: Log (work hours). Control group: Rejected DI applicants.

	100 percent DI vs. Rejected						Partial DI vs. Rejected					
	(1) Female	(2) Male	(3) Pooled	(4) Female	(5) Male	(6) Pooled	(7) Female	(8) Male	(9) Pooled	(10) Female	(11) Male	(12) Pooled
Age 18-61												
Diff-in-diff estimate	-0.029***	-0.020***	-0.025***	-0.000	0.008	0.003	-0.019***	-0.021***	-0.019***	-0.001	-0.001	-0.001
	(0.0067)	(0.0074)	(0.0053)	(0.0097)	(0.0074)	(0.0077)	(0.0059)	(0.0060)	(0.0045)	(0.0090)	(0.0062)	(0.0065)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[148589]	[116752]	[265341]	[148589]	[116752]	[265341]	[568513]	[219144]	[787657]	[568513]	[219144]	[787657]
Clusters	429	424	435	429	424	435	434	428	436	434	428	436
Age 18-29												
Diff-in-diff estimate	0.018	-0.064**	-0.024	0.024	-0.000	0.009	0.061*	-0.082**	-0.007	0.003	0.034	0.014
	(0.0321)	(0.0306)	(0.0223)	(0.0503)	(0.0342)	(0.0329)	(0.0340)	(0.0326)	(0.0261)	(0.0465)	(0.0353)	(0.0302)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[4494]	[6956]	[11450]	[4494]	[6956]	[11450]	[4289]	[3430]	[7719]	[4289]	[3430]	[7719]
Clusters	151	188	232	151	188	232	143	128	192	143	128	192
Age 30-49												
Diff-in-diff estimate	-0.043***	-0.027***	-0.037***	-0.012	-0.000	-0.006	-0.034***	-0.028***	-0.032***	-0.008	-0.009	-0.008
	(0.0099)	(0.0103)	(0.0073)	(0.0138)	(0.0100)	(0.0095)	(0.0081)	(0.0100)	(0.0060)	(0.0129)	(0.0096)	(0.0086)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[59381]	[44220]	[103601]	[59381]	[44220]	[103601]	[177710]	[69610]	[247320]	[177710]	[69610]	[247320]
Clusters	393	371	419	393	371	419	421	383	428	421	383	428
Age 50-61												
Diff-in-diff estimate	-0.016*	-0.005	-0.012*	0.007	0.016	0.010	-0.005	-0.002	-0.004	0.003	0.006	0.005
	(0.0081)	(0.0099)	(0.0065)	(0.0087)	(0.0105)	(0.0078)	(0.0071)	(0.0086)	(0.0058)	(0.0072)	(0.0086)	(0.0062)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[84714]	[65576]	[150290]	[84714]	[65576]	[150290]	[386514]	[146104]	[532618]	[386514]	[146104]	[532618]
Clusters	415	402	430	415	402	430	432	413	434	432	413	434

Intensive margin DD placebo estimates using individual and monthly fixed effects, with an observation window of +/- 12 months relative 1 January 2014. The first panel contains pooled regressions for all individuals 18-61 years of age, where each regression is run separately for women, men and both genders. The next three panels contain a subgroup analysis focusing on gender and age effects. The outcome variable is a log-transformed working hours variable and (standard errors) are clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01

Table 8: 2013/2014 Difference-in-difference placebo estimates: Percentage point change in labor market participation. Control group: Rejected DI applicants.

	All DI vs. Rejected					
	(1) Female	(2) Male	(3) Pooled	(4) Female	(5) Male	(6) Pooled
Age 18-61						
Diff-in-diff estimate	-0.003	-0.009***	-0.006***	0.005**	0.003	0.004**
	(0.0023)	(0.0026)	(0.0016)	(0.0023)	(0.0030)	(0.0018)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[2980578]	[2189766]	[5170344]	[2980578]	[2189766]	[5170344]
Clusters	436	440	440	436	440	440
Age 18-29						
Diff-in-diff estimate	-0.012	-0.025***	-0.019***	0.011	0.000	0.006
	(0.0108)	(0.0087)	(0.0071)	(0.0097)	(0.0086)	(0.0068)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[119873]	[154370]	[274243]	[119873]	[154370]	[274243]
Clusters	408	415	424	408	415	424
Age 30-49						
Diff-in-diff estimate	-0.003	-0.015***	-0.008***	0.007**	-0.002	0.003
	(0.0038)	(0.0038)	(0.0028)	(0.0032)	(0.0037)	(0.0024)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[1011413]	[792568]	[1803981]	[1011413]	[792568]	[1803981]
Clusters	435	435	438	435	435	438
Age 50-61						
Diff-in-diff estimate	0.003	0.004	0.003	0.003	0.007	0.005*
	(0.0031)	(0.0050)	(0.0030)	(0.0035)	(0.0043)	(0.0029)
Trend adjustment	No	No	No	Yes	Yes	Yes
Observations	[1849292]	[1242828]	[3092120]	[1849292]	[1242828]	[3092120]
Clusters	436	438	438	436	438	438

Extensive margin DD estimates using individual and monthly fixed effects, with an observation window of +/- 12 months relative 1 January 2014. The first panel contains pooled regressions for all individuals 18-61 years of age, where each regression is run separately for women, men and both genders. The next three panels contain a subgroup analysis focusing on gender and age effects. The outcome variable is a dummy variable for working more than four hours and (standard errors) are clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01

Table 9: 2013/2014 Difference-in-difference placebo estimates: Percentage point change in labor market participation. Control group: Rejected DI applicants.

	100 percent DI vs. Rejected						Partial DI vs. Rejected					
	(1) Female	(2) Male	(3) Pooled	(4) Female	(5) Male	(6) Pooled	(7) Female	(8) Male	(9) Pooled	(10) Female	(11) Male	(12) Pooled
Age 18-61												
Diff-in-diff estimate	0.006**	-0.005**	0.001	0.006***	0.003	0.005***	-0.017***	-0.020***	-0.020***	0.003	0.003	0.003
	(0.0025)	(0.0026)	(0.0017)	(0.0023)	(0.0031)	(0.0017)	(0.0026)	(0.0026)	(0.0018)	(0.0027)	(0.0032)	(0.0022)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[2310505]	[1911639]	[4222144]	[2310505]	[1911639]	[4222144]	[801271]	[385503]	[1186774]	[801271]	[385503]	[1186774]
Clusters	436	439	439	436	439	439	435	438	438	435	438	438
Age 18-29												
Diff-in-diff estimate	-0.011	-0.025***	-0.018***	0.011	0.000	0.005	-0.053*	-0.022	-0.042**	0.032	-0.001	0.019
	(0.0108)	(0.0087)	(0.0070)	(0.0100)	(0.0087)	(0.0069)	(0.0271)	(0.0249)	(0.0205)	(0.0230)	(0.0187)	(0.0156)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[115260]	[151469]	[266729]	[115260]	[151469]	[266729]	[14287]	[13068]	[27355]	[14287]	[13068]	[27355]
Clusters	404	414	423	404	414	423	242	236	300	242	236	300
Age 30-49												
Diff-in-diff estimate	0.002	-0.014***	-0.005*	0.007**	-0.001	0.003	-0.015***	-0.024***	-0.020***	0.008*	-0.002	0.004
	(0.0039)	(0.0038)	(0.0028)	(0.0032)	(0.0038)	(0.0024)	(0.0046)	(0.0047)	(0.0034)	(0.0041)	(0.0042)	(0.0031)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[808394]	[708921]	[1517315]	[808394]	[708921]	[1517315]	[269843]	[137141]	[406984]	[269843]	[137141]	[406984]
Clusters	435	434	437	435	434	437	428	417	432	428	417	432
Age 50-61												
Diff-in-diff estimate	0.015***	0.009*	0.012***	0.005	0.007	0.006**	-0.010***	-0.007	-0.010***	0.000	0.006	0.002
	(0.0030)	(0.0050)	(0.0029)	(0.0035)	(0.0043)	(0.0029)	(0.0035)	(0.0048)	(0.0033)	(0.0037)	(0.0047)	(0.0031)
Trend adjustment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Observations	[1386851]	[1051249]	[2438100]	[1386851]	[1051249]	[2438100]	[517141]	[235294]	[752435]	[517141]	[235294]	[752435]
Clusters	436	438	438	436	438	438	434	432	436	434	432	436

Extensive margin DD placebo estimates using individual and monthly fixed effects, with an observation window of +/- 12 months relative to 1 January 2014. The first panel contains pooled regressions for all individuals 18–61 years of age, where each regression is run separately for women, men, and both genders. The next three panels contain subgroup analyses focusing on gender and age effects. The outcome variable is a dummy variable for working more than four hours and standard errors are clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01.

Appendix B

$$B^{old} = \begin{cases} & if & w \leq SGA_1 \\ B^1(1-t_p) & where & SGA = \$11\,856\,p.a \\ & and & t_p = pension\,tax \\ \\ & if & SGA_1 < w \leq SGA_2 \\ (B^1\text{-}\Theta)(1\text{-}t_p) & where & \Theta = \frac{w}{IBD}\,B^1 \\ & and & \frac{w}{IBD}\,is\,rounded\,to\,the\,closest\,5\,percent \\ & and & IBD = Adjusted\,income\,before\,DI \\ \\ (B^1-\Theta-\gamma)(1-t_p) & if & SGA_2 < w \leq SGA_3 \\ & where & \gamma = \frac{w}{IBD}(B^1-\theta) \\ & and & \frac{w}{IBD}\,is\,rounded\,to\,the\,closest\,5\,percent \\ \\ ... & & \\ 0 & if & w + B^1 \geq IBD \end{cases}$$

The pre-reform DI benefits (B^1) are described above as a function of earnings (w), for an individual that was evaluated to have lost 100% of the ability to earn an income. Under the pre-reform system, the individual receives 100% DI benefits (B^1) minus pension tax (t_p), as long as income (w) stays below the yearly significant gainful activity (SGA) threshold. When the income *(w)* exceeds the SGA_1 threshold, the disability level is reduced by the size of the yearly income relative to the pre-reform estimated income ($\frac{w}{IBD}$), rounded to the nearest five percent. The DI benefits reduce concurrently with the disability level; a wage equaling 28% of adjusted income before disability insurance receipt (IBD), which exceeds the SGA_1 threshold, would reduce the disability level to 70% prior to 1 January 2015. Breaking the SGA_1 threshold by a marginal amount would therefore result in a 30% reduction in the DI benefits. This meant DI beneficiaries had few incentives to earn an income above the first SGA threshold prior to the 2015 reform; a significant increase in the labor supply would be necessary to offset the large reduction in the DI benefits resulting from the marginal increase in the yearly salary. The discontinuity pattern in the budget restriction would repeat at the SGA_2 threshold given to the individual now receiving 70% partial DI benefits. The old DI system also had an absolute threshold, where the sum of DI benefits and yearly income from labor (B^1+w) could not exceed income before the point of disability (IBD).

$$B^{new} = \begin{cases} B^0(1-t_i) & \begin{array}{c} if \\ where \\ and \end{array} & \begin{array}{c} w \leq SGA \\ SGA = \$4742\, p.a \\ t_i = income\, tax \end{array} \\ (B^0 - r(w-SGA))(1-t_i) & \begin{array}{c} if \\ where \end{array} & \begin{array}{c} SGA < w + B^0 \leq IBD * 0,8 \\ r = \frac{B^0}{IBD} \end{array} \\ 0 & \begin{array}{c} if \\ where \end{array} & \begin{array}{c} w \geq IBD * 0,8 \\ IBD = Adjusted\, income\, before\, DI \end{array} \end{cases}$$

The DI system is more transparent for DI beneficiaries following the 2015 reform. A 100% DI beneficiary receives 100% DI benefits (B^0) minus income tax, as long as the income from labor (w) stays below the new \$4,742 SGA threshold. However, the SGA threshold is \$8,042 until 2019 for all individuals awarded DI benefits prior to the implementation of the 2015 reform. After tax, the pre-reform level of DI benefits (B^1(1-t$_p$)) should approximately equal the post-reform benefit level (B^0(1-t_i)) for all DI beneficiaries.

As income from labor *(w)* exceeds the new \$4,742 SGA threshold, the DI benefits are gradually reduced at the real rate of income replacement ($\frac{B^0}{IBD}$) after the DI reform. This reduction model continues until the wage *(w)* from labor exceeds 80% of income before the point of disability (IBD). At this 80% level, DI benefits are reduced to zero, but the beneficiary retains the right to start a benefit claim if his or her health deteriorates after some time in the labor force.

Appendix C

The 2015 DI reform was implemented on the same day that the rules for reporting labor market information to the government were changed. This reporting change affects the labor market data that I use in the analysis. Prior to January 2015, employers sent five reports separately each month to the Labor and Welfare Administration, Statistics Norway and the Norwegian Tax Administration. Starting in January 2015, employers send one report to the government, which then distributes the report information among the relevant government agencies. The reporting change therefore reduced the monthly reporting from employers to the government from five reports to one.

Prior to the reporting change, employers reported individuals that worked more than four hours on average per week over more than 14 days. Since the reporting change, they report everyone that makes more than $134 per year. Because more individuals are reported as working after the reporting change, I have cleaned the data to make them similar before and after January 2015. I removed observations of individuals that work less than four hours per week per month, because these individuals were not reported as working prior to January 2015. If an employee worked parts of a month, the labor supply for each employer is averaged and summed for each individual to reflect the total monthly labor supply. I have also used a correction that is used by Statistics Norway, where individuals do not work more than 45 hours per week for one employer, and no more than 60 hours in total for all employers. This correction affects some obvious reporting errors, where some individuals work an unrealistic number of hours per week.

Department of Economics
University of Bergen
PO BOX 7800
5020 Bergen
Visitor address: Fosswinckels gate 14
Phone: +47 5558 9200
www.uib.no/econ/