



Income and substitution effects of a disability insurance reform[☆]

Eva Deuchert^a, Beatrix Eugster^{b,*}

^aUniversity of Fribourg, Switzerland

^bCDI-HSG, University of St. Gallen, Rosenbergstrasse 51, St. Gallen CH-9000, Switzerland

ARTICLE INFO

Article history:

Received 28 June 2017

Received in revised form 5 December 2018

Accepted 5 December 2018

Available online xxx

JEL classification:

C30

I13

J01

Keywords:

Disability insurance

Work disincentives

Income and substitution effects

Partial benefit system

ABSTRACT

Disability insurance (DI) systems are widely criticized for their inherent work disincentives. This paper evaluates the effects of a Swiss DI reform that aims to lower DI benefits for a group of existing DI beneficiaries and introduces an additional level to the DI benefit schedule. The reform has only modest effects on earnings and employment but increases the disability degree of those threatened by a DI benefit decline. We estimate bounds on the income and substitution effects by employing the principal stratification framework. The income effect is quantitatively important, whereas the substitution effect is smaller and has bounds that include zero. The evidence suggests that caseworkers helped the insured with low labor market attachment to maintain a full DI benefit.

© 2018 Published by Elsevier B.V.

1. Introduction

The large number of people with disabilities, their low labor market attachment and their high dependency on social assistance create considerable costs to society (OECD, 2010). Many countries are thus forced to reform their disability insurance (DI) systems. One of the prime problems to be solved is the work disincentive effects of DI systems. Because DI beneficiaries fear losing a significant portion of their benefits if labor supply exceeds certain thresholds – so called “cash cliffs” – they do not increase employment above this level (substitution effect). Furthermore, DI benefits increase non-earned income, which reduces employment if people prefer leisure over labor (income effect). Although these two mechanisms are well understood in theory, identifying income and substitution effects empirically is challenging. Individual reactions to changes in the

benefit schedule usually reflect both mechanisms jointly and are therefore not informative of either effect.

This paper presents novel insights into the importance of the income and substitution effects by evaluating a reform of the Swiss DI system and employing a principal stratification approach. The Swiss system insures partial disability, where beneficiaries can work and claim DI benefits simultaneously. The level of DI benefits is a step-wise function of disability degree, which is assessed by the DI and represents the presumed earnings loss due to the disability (in percent). In January 2004, Switzerland further graduated the DI benefit system and introduced a three-quarter DI benefit, in addition to the existing one-quarter DI benefit, one-half DI benefit, and full DI benefit. The reform led to a substantial loss in DI benefits for a subset of beneficiaries and imposed a new earnings threshold for the full benefit. The theoretical effects are conflicting. The income effect is likely to increase labor supply because the loss in DI benefits must be compensated by an increase in earnings. The substitution effect, as is called the bunching that arises at nudges of the benefit schedule (Autor and Duggan, 2007), is likely to reduce incentives to work because a reduction in earnings signals an increase in disability degree and therefore can lead to a preservation of the full DI benefit.

We first evaluate the average effect of the reform on employment and earnings using a local difference-in-differences approach taking advantage of the fact that only individuals born after

[☆] We are grateful for the comments from Rafael Lalive, Helge Liebert, Judit Vall Castello, the conference participants at Basel and Nyborg, and the seminar participants at the Universities of St. Gallen, Fribourg, Lugano, Lausanne, University College Dublin, and KOF Zurich. All remaining errors are our own. We thank the Federal Social Insurance Office (FSIO) for the data provision and support. We acknowledge the financing from the Swiss National Science Foundation (grant no. 100018_143317/1).

* Corresponding author.

E-mail address: beatrix.eugster@unisg.ch (B. Eugster).

December 31, 1953, were fully subjected to the reform, whereas individuals born before December 31, 1953, were exempted from the benefit cut. We then use the principal stratification framework (Frangakis and Rubin, 2002) to decompose the average effect and provide bounds for the income and substitution effects. The resulting bounds can be sequentially tightened by adding revealed preference restrictions motivated by a simple static labor supply model.

We find a modest average effect of 2.3 percentage points on employment and no effect on earnings. Decomposing the average effect, we find informative bounds for the income effect, where the reform increased employment for individuals actually losing 25% of their DI benefits by 9 to 20 percentage points and increased earnings by 136 to 3135 CHF, which is up to 50% of the mean pre-reform earnings. The share of individuals complying to the reform and thus experiencing a decline in DI benefits is, however, small (25%). The majority of individuals kept a full DI benefit (75%), but the results document that a reduction in labor supply is not to blame: the bounds on the substitution effect are far smaller than those on the income effect and suggest only a small labor supply reaction to the reform of individuals who kept a full DI benefit. Individuals had thus other possibilities to signal a higher disability degree to keep a full DI benefit. This result suggests that caseworkers helped the affected individuals by lifting their disability degree over the threshold of 70% to ensure they were still eligible for a full DI benefit. We indeed find an immediate and persistent increase in the disability degree of approximately 3 percentage points, which is driven by individuals who had earnings below the threshold before the reform.

A relatively abundant literature on work disincentives imposed by DI exists, in which implicit or explicit changes to the budget constraint are used to derive structural parameters of labor supply (see Bound and Burkhauser, 1999 for a review). The existing literature exploits reforms on the generosity of the DI system (Campolieti and Riddell, 2012; Gruber, 2000; Kauer, 2014; Kostol and Mogstad, 2014; Marie and Vall Castello, 2012; Schimmel et al., 2011; Weathers and Hemmeter, 2011) or eligibility criteria (Autor and Duggan, 2007; Borghans et al., 2014; Karlström et al., 2008; Moore, 2014; Staubli, 2011) or compares the labor supply of accepted and rejected DI applicants (Bound, 1989; Chen and van der Klaauw, 2008; French and Song, 2014; Maestas et al., 2013; von Wachter et al., 2011). Overall, the findings suggest that the DI system imposes work disincentives on some individuals in addition to providing income to individuals who are at need.

Our study differs in three important ways from the previous literature. First, to the best of our knowledge, no literature on the performance of a partial DI benefit system exists, even though many countries (such as France, Germany, Netherlands, Spain, Sweden and Switzerland) already rely on partial DI systems and there is also increasing interest in the US to provide partial income support (Autor and Duggan, 2010).¹ Second, currently, only three studies have estimated the income effects of DI. Two studies use the fact that non-work-contingent systems do not create substitution effects (Autor and Duggan, 2007; Marie and Vall Castello, 2012). In the third study, Gelber et al. (2017) estimate local income effects at the upper bend point of the SSDI benefit schedule in the US and find that a \$1 increase in benefits reduces earnings by \$0.2. Their estimates on employment effects are less conclusive. A better understanding of the importance of income and substitution effects is crucial for designing effective DI systems. Third, we develop a novel empirical framework motivated by a labor market model that allows derivation of bounds on these conflicting effects. This methodology is related to Kline and Tartari (2016), who study the bounds of labor supply

responses to the US job first program within a revealed preferences framework.

The rest of this paper proceeds as follows. Section 2 provides details of the Swiss DI Act. Section 3 outlines the expected effects of the reform. Section 4 discusses the empirical identification strategy. Section 5 presents the data and descriptive statistics, and Section 6 gives the results and provides detailed evidence for the validity of the parallel trend assumption. Section 7 discusses potential caseworker reactions to the reform, and Section 8 concludes.

2. Swiss Disability Insurance Act

In Switzerland, the mandatory public DI insures individuals against partial or full loss of the ability to work due to impaired health. DI is permanent, i.e., an individual can claim disability benefits for as long as the health condition remains unchanged. Overall, the system is very generous. The public DI system, together with the occupational pension scheme, guarantees replacement rates of at least 60% of previous earnings in the case of a full DI benefit, and in most cases the benefit is much higher.

The Swiss DI system allows for partial disability. DI benefits are a stepwise function of the disability degree. In contrast to other countries, the disability degree is not a function of the physical work capacity but denotes the presumed earnings loss due to the disability (in percent) and is determined by the DI office. Specifically, the disability degree is assessed by caseworkers in the following manner:

$$\text{disability degree (dd)} = \left(1 - \frac{\text{potential earnings with disability}}{\text{potential earnings without disability}}\right) * 100\%$$

The eligibility for DI benefits, as well as the disability degree, is determined by a DI caseworker based on medical documentation and previous earnings records. Typically, *potential earnings without disability* are predicted on the basis of the individual's earnings before the onset of the disability. The potential earnings with disability are more difficult to determine. As long as the caseworker concludes that the person exhausted full work capacity, *potential earnings with disability* is set on the basis of the individual's earnings during disability. If the caseworker concludes that the person has idle work capacity, for example, because medical records suggest that the person could work a larger number of hours, potential earnings can be fixed based on the assumed work capacity and official wage indices. As a consequence, many insured individuals are not employed but still have a disability degree less than 100%. Moreover, the disability degree is not a time-invariant number, and regular reassessments are initiated by either the DI or the insured. If, for example, employment changes, the beneficiary is legally obligated to declare higher income to the DI system to reassess the DI status. Reassessment can also be requested when health deteriorates.

In this paper, we evaluate whether the stepwise DI system imposes work disincentives. Table 1 gives an overview of the benefit schedule in place before January 1, 2004. Individuals with a disability degree lower than 40% receive no DI benefits, those with a disability degree between 40% and 49% receive a one-quarter DI benefit, and those with a disability degree between 50% and 66% receive a one-half DI benefit. Finally, a minimum disability degree of 67% makes individuals eligible for a full DI benefit.

We study a policy change that was implemented as part of the fourth revision of the Swiss DI system since January 1, 2004. The reform introduced the three-quarter DI benefit for individuals with a disability degree between 60 and 70%, in addition to the already existing one-quarter, one-half, and full DI benefits (see Table 1). Therefore, individuals with a disability degree between 60

¹ An exception is Borghans et al. (2014) who study social support substitution from a reform of a partial DI system in the Netherlands.

Table 1
The partial DI system (before/after the reform).

DI benefit level	Before Jan 1, 2004	Since Jan 1, 2004
Full	67% ≤ dd	70% ≤ dd
Three-quarter	None	60% ≤ dd < 70%
One-half	50% ≤ dd < 67%	50% ≤ dd < 60%
One-quarter	40% ≤ dd < 50%	No change
No DI benefits	dd < 40%	No change

Note: dd stands for disability degree and represents the potential earnings loss due to disability (in percentage) compared to earnings potential without a disability.

and 66% are lifted from a one-half to a three-quarter DI benefit due to the reform, whereas individuals with a disability degree between 67 and 69% lose a quarter of their DI benefits if their disability degree remains unchanged. Furthermore, the change in the benefit schedule introduces two new disability degree notches, at 60% for a three-quarter DI benefit and at 70% for a full DI benefit. In this paper we evaluate behavioral responses to a potential reduction in DI spendings by focusing on individuals with an initial disability degree between 67 and 69%. For this subgroup, the reform provides a suitable control group because individuals who were older than 50 years in January 2004 were exempted from the DI benefit cut, thus enabling an evaluation of the effect of the reform using quasi-experimental empirical methods.²

In this context, concerns about anticipation effects may arise. The policy change under study was implemented as part of the fourth revision of the Swiss DI Act, which was already planned in the late 1990s. However, the introduction of the three-quarter DI benefit was first discussed in 2002, passed the parliament in 2003, and became effective on January 1, 2004. Thus, limited time was available for adjustments before the reform was implemented. We will further discuss anticipation effects in Section 6.3.

3. Predicted effects in a revealed preferences framework

The expected effects of the reform can be predicted in a simple static labor supply model: Total disposable income y consists of earnings Y and disability benefits B . The DI suffers from an asymmetric information problem because DI caseworkers cannot observe the true disability degree. Assume that the caseworker sets potential earnings without disability equal to the last earnings before the onset of the disability and that potential earnings with disability are equal to observed current earnings. The resulting thresholds in current earnings imposed by the step-wise DI system are thus unique for each individual and depend on their earnings before the onset of disability. Individuals can signal a higher disability degree by choosing a lower employment level. Since our analysis focuses on individuals with disability degrees prior to the reform of between 67 and 69%, for simplicity, we assume a single notch:

$$y = \begin{cases} Y + B & \text{if } Y \leq \pi \\ Y + \alpha B & \text{if } Y > \pi \end{cases}$$

where $(1 - \alpha)B$ is the size of the notch, and π is the earnings threshold for a full DI benefit.

Fig. 1 shows the budget constraints before and after the reform. Note that we label earnings on the abscissa in terms of a percentage relative to pre-disability earnings to simplify the comparability with other settings, where fixed absolute earnings thresholds (such as the

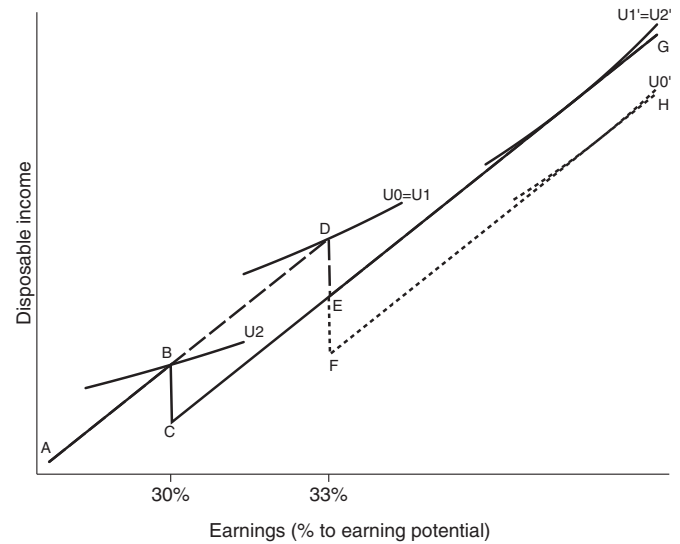


Fig. 1. Expected effect of the reform. Note: The line ADFH is the budget constraint for all individuals before the DI reform. The line ABCG is the budget constraint for individuals eligible for DI benefits and younger than 50 years in January 2004. The line ADEG is the budget constraint for individuals older than 50 years in January 2004. Disposable income is earnings plus DI benefits. Earnings are in percentage of earnings potential (pre-disability earnings). U_i denotes utility level if the person received a full DI benefit, and U_i' is the utility level if the person receives a partial DI benefit. The subscript i denotes the utility levels under different reforms (i.e., 0 without the reform, 1 for individuals older than 50 years, and 2 for individuals younger than 50 years).

SGA threshold) are modified. Therefore, a disability degree of 67% – the threshold for full DI benefits before the reform – would translate to an income of 33% relative to the pre-disability earnings. Similarly, the threshold of 70% after the reform translates to an income of 30% relative to pre-disability earnings. Before the reform, the budget constraint is identical for all individuals in relative terms (ADFH). Individuals have the choice to sacrifice one-half of their DI benefits and increase their earnings above the 33% threshold or to reduce labor supply to less than 33% of their pre-disability earnings and thus receive the full DI benefit. The individual chooses a full DI benefit if the utility U of an employment level of 33% or less (U_0) is higher than the utility from higher employment levels (U_0').

After the reform, the shape of the budget constraint depends on age. Individuals who are 50 years and older experience a parallel shift in the budget constraint for all earnings above the 33% threshold. The cut-off threshold for a full DI benefit, however, remains constant at 33% since these individuals are exempted from the benefit cut. Their budget constraint is thus equal to ADEG. These individuals now sacrifice only one-fourth of their DI benefit when their employment exceeds 33%. Individuals who are younger than 50 years additionally experience a shift in the earnings threshold for a full DI benefit from 33% to 30%. Their new budget constraint is ABCG.

The empirical part of this paper compares labor outcomes of individuals who were fully subjected to the reform to those of individuals who were exempted from the benefit cut because of their age. Therefore, we compare the difference between the solid and the dashed line in Fig. 1 rather than evaluate the full reform, which corresponds to the difference between the dotted line and the solid or the dashed line.

Individuals may react to the threshold change in different ways. We borrow from the nomenclature in Angrist et al. (1996) and analyze the consequence of the threshold change for *compliers*, *never-takers*, and *always-takers*. Compliers reduce DI benefits to a partial DI benefit when subjected to the new threshold but maintain a full DI benefit when subjected to the old threshold. Because DI benefits decline, individuals may increase employment due to the income

² Further features of the reform were the abolition of additional spousal benefits for newly entering beneficiaries (Kauer, 2014), hardship benefits, an increase in the helpless allowance for individuals with special care needs and the development of medical expertise for the evaluation of DI benefit appraisals via the introduction of regional screening centers (Liebert, 2016). These parts of the reform affect either only inflow or all insured individuals equally and are not tied to any age threshold.

effect. In terms of our simple model in Fig. 1, this result occurs if $U_1 > U_2' > U_2$. Suppose utility is just a function of disposable income; the person would always choose employment at the highest notch. Denote full time with X and assume that full time income is not affected by disability (in other words, disability does not affect profession or wage, only the number of hours worked). Disposable income at U_1 is equal to $B + 0.33X$ (with earnings $Y = 0.33X$), disposable income at U_2' is $3/4B + 0.4X$, and disposable income at U_2 is $B + 0.3X$. To be a complier, the replacement rate (i.e., B/X) must be between 28 and 40%. Given that the Swiss DI benefit system typically guarantees much higher replacement rates, the share of compliers is likely to be small. One should note, however, that our simple model disregards adjustment costs in labor contracts. Rather than bunching at the next notch, it is thus also possible that individuals need to reduce employment or earnings by a larger extent or stop working altogether to maintain a full DI benefit. If this is the case, the share of compliers could be far larger, as disposable income at U_2 is less than $B + 0.3X$.

Never-takers do not reduce DI benefits, neither with the old nor the new notch. To keep a full DI benefit, however, they may need to signal a deterioration in health. If their earnings are above the new threshold, these individuals are forced to reduce earnings. For never-takers, the preference structure is thus $U_1 > U_2 > U_2'$. In the case that utility is a function of only disposable income, never-takers have replacement rates higher than 40% (or higher when adjustment costs are relevant). Given the high replacement rates in the Swiss DI system, a large fraction of our sample may belong to the group of never-takers. However, our model assumes that asymmetric information leads to a situation where caseworkers set potential earnings with disability equal to actual earnings when predicting the disability degree. In reality, this scenario may not be the case since caseworkers can fix higher potential earnings in case they have evidence that the person does not exploit their full earnings capacity. This evidence could be based on medical records or physicians' assessments suggesting that a person could still work a certain number of hours or work in certain jobs. In this case, reducing earnings below the income threshold is a necessary but not sufficient condition for keeping a full DI benefit. Caseworkers have the discretion to keep the disability degree unchanged even though actual earnings fall below the income threshold. Individuals with high replacement rates are therefore not necessarily never-takers but can be compliers (not by their own choice, but by the choice of the caseworker).

Finally, there may be always-takers who reduce DI benefits regardless of whether they are subjected to the new or old thresholds. Their utility structure is equal to $U_2' > U_1$ and $U_2' > U_2$, and they are likely to be individuals with relatively low replacement rates (less than 28% in the case that utility is a function of disposable income only). Note that the change in threshold has no effect on behavior since $U_2' = U_1'$.

Interestingly, the model also predicts that defiers do not exist. Defiers increase incomes above the income threshold of 33% and thus reduce DI benefits to three-quarter benefits in the case that the reform is not in place (i.e., $U_2' > U_1$) but restrain from doing so when the income threshold is 30% ($U_2 > U_2'$). This scenario is not possible, as under a well-behaved utility structure, it must hold that $U_1 > U_2$.

Two aspects of the definition of substitution and income effects as outlined above are non-standard. First, the substitution effect differs from the standard concept because it does not stem from a change in relative prices of consumption and leisure. Rather, it denotes the bunching that arises at so-called "cash cliffs" because DI beneficiaries fear losing a significant part of their benefits if labor supply exceeds these thresholds. The relevant literature calls this effect the substitution effect (Autor and Duggan, 2007; Marie and Vall Castello, 2012), even though this effect describes a corner solution rather than the change in the relative price level. Second, the measures of income and substitution effects are not defined within

individuals but rather across groups of individuals characterized by their reaction to the change in the DI benefit schedule. While every individual's preferences determine the income and substitution effects for each individual, we are able to identify the income effect only for those individuals where the income effect actually dominates the substitution effect, and vice versa.

In the following, we are interested in analyzing the average effect of the difference in thresholds on labor market outcomes, as well as the strata-specific effects that correspond to the income and substitution effects.

4. Identification

4.1. Average effect of the reform

We first estimate the average effect of the reform on employment, earnings, and disability degree. Our sample consists of individuals who had an initial disability degree prior to the reform of between 67 and 69% and thus were subjected to a benefit cut and a new threshold for a full DI benefit. We apply a difference-in-differences (DiD) framework and make use of the fact that individuals aged 50 years or older in January 2004 are exempt from the benefit cut and thus serve as a control group.³

$Z_i = 1$ denotes the age cohort being exposed to the benefit cut (younger than 50 years in January 2004), and $Z_i = 0$ denotes the age cohort not exposed (older than 50 years in January 2004). $Y_{it}(Z_i)$ denotes potential outcome, where for each individual, we can observe only $Y_{it} = Z_i Y_{it}(1) + (1 - Z_i) Y_{it}(0)$. That is, each individual is observed in only one of the two potential states, and the individual effect of the reform $TE_{it} = Y_{it}(1) - Y_{it}(0)$ cannot be identified. The simple comparison between average outcomes after the reform in the treatment and the control groups is biased by non-random treatment allocation (selection bias) because of age trends in health, employment, and earnings. We employ a DiD strategy to correct for this bias and estimate the average effect of the reform as follows:⁴

$$E[Y_t(1) - Y_t(0)|Z = 1] = E[Y_t - Y_0|Z = 1] - E[Y_t - Y_0|Z = 0]$$

where the subscript t denotes the years after the introduction of the reform, and the subscript 0 refers to the last period before the reform (i.e., 2003). The main identifying assumption underlying the DiD strategy is that treated and control individuals exhibit a parallel trend in outcomes before the reform. One should note that both groups, namely, treated and controlled, are exposed to the same aspect of the reform, i.e., the introduction of the three-quarter DI benefit rather than the one-half DI benefit for disability degrees between 60 and 66%. The identifying assumption is thus that younger and older individuals react in the same way to introducing the three-quarter DI benefit given that the cut-off level for a full DI benefit was unchanged at 67%. This relates to our simple model (Fig. 1) in the following way. Both age cohorts are exposed to the same reform that corresponds to the change of the dotted line (FH) to the solid line

³ Other methods could be used to estimate the effect. We could employ an RDD-design. However, given that our sample is already restricted to individuals with initial disability degrees between 67 and 69%, our sample is relatively small. We employ this method for consistency/sensitivity analysis. Alternatively, previous literature has exploited bunching to estimate behavioral responses to kinks or discontinuities in tax and benefit schedules (i.e. Brown, 2013; Ruh and Staubli, 2015; Saez, 2010). However, this approach relies on the assumption that the counterfactual distribution of the disability degree is continuous or smooth. This assumption is clearly violated in the current context, where we observe strong bunching at decimal disability degrees that are not associated with payout thresholds (such as 70%, 80%, and 100%). Our identification strategy is a classic control group design and is thus not biased by any discontinuity in the distribution that is not caused by the effect of interest.

⁴ From now on, the subscript i will be omitted when possible.

(EG). Only individuals with $U1' > U1$ react to this part of the reform and are expected to increase labor supply. We assume that the share of individuals with $U1' > U1$ is the same in both age cohorts. We show supporting evidence for the parallel trend assumption in Section 6.3.

Since we have panel data, we implement this method via regression on the data in first differences

$$\Delta Y_{it} = \alpha + \beta Z_i + u_i, \quad (1)$$

where $\Delta Y_{it} = Y_{it} - Y_{i0}$, Y_{it} is the observed outcome for the years following the reform, Y_{i0} is the observed outcome for the year 2003, and β is the average effect of the reform.⁵

4.2. Income and substitution effects

The average effect of the reform on employment and earnings is composed of conflicting income and substitution effects. We use the principal stratification framework (Frangakis and Rubin, 2002) to decompose the average effect into group-specific effects for different strata of individuals. In Section 3 we demonstrated that the individual labor supply reaction is due to either the income or the substitution effect but that both effects are not relevant for the same individuals. Therefore, causal effects for the different principal strata have a straightforward economic interpretation as income and substitution effects.

The following section outlines our empirical strategy to identify causal effects for different strata. The exact formulation of the resulting bounds is summarized in Table 2, and the mathematical proofs are provided in the appendix.

Denote potential partial DI benefit receipt with an indicator $D_t(Z) \in \{0, 1\}$, which is equal to one if a person receives a partial (i.e., three-quarter) DI benefit and equal to zero if the person receives a full DI benefit. The indicators for potential partial DI benefit receipt and exposure to the reform are both binary, allowing decomposition of the full population into four different strata ($S_t = s$). For two groups, the DI benefit level does not change due to the reform: never-takers ($S_t = nt$) never reduce DI benefits to partial benefits ($D_t(1) = 0, D_t(0) = 0$), while always-takers ($S_t = at$) always receive a partial DI benefit ($D_t(1) = 1, D_t(0) = 1$). The remaining two strata react to the reform. Compliers ($S_t = c$) reduce their DI benefit level as a result of exposure to the reform ($D_t(1) = 1, D_t(0) = 0$), while defiers ($S_t = d$) show the exact inverse reaction ($D_t(1) = 0, D_t(0) = 1$).

The principal effect with respect to a principal stratum is defined as the comparison of potential outcomes within a stratum. The average effect of the reform can be decomposed to:

$$\begin{aligned} E[Y_t(1) - Y_t(0)|Z = 1] &= E[Y_t(1) - Y_t(0)|Z = 1, S_t = nt]Pr(S_t = nt|Z = 1) \\ &+ E[Y_t(1) - Y_t(0)|Z = 1, S_t = c]Pr(S_t = c|Z = 1) \\ &+ E[Y_t(1) - Y_t(0)|Z = 1, S_t = d]Pr(S_t = d|Z = 1) \\ &+ E[Y_t(1) - Y_t(0)|Z = 1, S_t = at]Pr(S_t = at|Z = 1) \end{aligned} \quad (2)$$

This decomposition brings the econometric model to the predictions of the simple labor supply model: to keep full DI benefits after the reform, never-takers need to increase their disability degree by signaling lower earnings potential. This may come at the cost of reducing labor supply so that earnings fall below the new earnings threshold. We therefore expect a negative earnings effect for never-takers due to the substitution effect $SE_t = E[Y_t(1) - Y_t(0)|Z = 1, S_t = nt] \leq 0$. Note that this result is at odds with the standard instrumental variable setting, where never-takers are assumed not to react to the instrument (exclusion restriction). Compliers, by contrast, accept the benefit cut and potentially increase earnings due to the income effect $IE_t = E[Y_t(1) - Y_t(0)|Z = 1, S_t = c] \geq 0$. Estimating the principal effects thus yields the effects of interest for the two principal groups affected by the reform.

Analogously to the estimation of the average effect of the reform, we employ a difference-in-differences specification. Under the assumption of parallel trends within each strata (**Assumption 1.a**), we can estimate the average treatment effects on the treated group within each strata if the strata were observed. However, strata depend on latent variables and can therefore not be directly observed. Under the additional assumption that exposure to the reform is independent of potential DI benefit receipt $D_{i,t}(1), D_{i,t}(0) \perp Z_i$ (**Assumption 1.b**), the strata proportions are identical in the treated and control groups. Assumption 1.b implies that age has no influence on the strata proportions. This is a strong assumption because deteriorating health or adverse labor market conditions for elderly individuals could influence the probability of being a never-taker vs. a complier. We therefore investigate this assumption in Section 6.3. Note that assumptions 1.a and 1.b impose parallel trends in the full sample, which were needed to identify the average effect of the reform.

To partially identify substitution and income effects, we rely on an additional set of assumptions that is backed by our simple labor supply model. Individual level monotonicity $D_{i,t}(1) \geq D_{i,t}(0)$ (**Assumption 2**) assures that defiers do not exist. Nobody would decide to increase labor supply such as to receive only a partial DI benefit in the absence of the reform but not increase or rather decrease labor supply to keep a full DI benefit if affected by the reform. Individuals who choose a three-quarter DI benefit over a full DI benefit in the absence of the reform reveal that their utility from expanding employment and lowering DI benefits is higher than the utility from bunching earnings at the old threshold, which is the necessary requirement to keep the full DI benefit without the reform. Together, assumptions 1.b and 2 allow the identification of strata proportions p_s :

$$\begin{aligned} p_{nt} &= Pr(D_t = 0|Z = 1) \\ p_{at} &= Pr(D_t = 1|Z = 0) \\ p_c &= Pr(D_t = 1|Z = 1) - Pr(D_t = 1|Z = 0) \end{aligned}$$

These rather standard assumptions allow constructing bounds for the principal strata effects (Eq. (2)). Let us first consider the effect for never-takers, which we interpret as the substitution effect of the reform. Since defiers are assumed not to exist (Assumption 2), all treated individuals who keep a full DI benefit are considered to be never-takers. We thus directly observe the first component of the first-difference estimator, i.e., $E[\Delta Y_t|Z = 1, S_t = nt] = E[\Delta Y_t|Z = 1, D_t = 0] = E[\Delta Y_t^{10}]$. However, never-takers and compliers are observationally equivalent in the observed group with $Z = 0$ and $D_t = 0$. Therefore, the remaining component of the first-difference estimator $E[\Delta Y_t|Z = 0, S_t = nt]$ cannot be observed directly. However, the relative group size of never-takers in the observed subgroup $Z = 0$ and $D_t = 0$ can be estimated (i.e., $pr_{nt} = \frac{p_{nt}}{p_{nt} + p_c}$). To estimate

⁵ Note that this estimation procedure cannot take into account endogenous outflow. From 2003 to 2004, outflow in the treated and control groups was low and comparable (2.7% vs. 2.2%). The treatment and the included covariates cannot significantly predict outflow, except for the gender dummy. Our analysis adds control variables, including gender.

Table 2
Bounds for income and substitution effect.

Ass.	Lower bound	Upper bound
Compliers: $S_t = c$		
1.a, 1.b, 2	$E[\Delta Y_t^{11} \Delta Y_t^{11} \leq \Delta y_{t,prc}^{11}] - E[\Delta Y_t^{00} \Delta Y_t^{00} > \Delta y_{t,prnt}^{00}]$	$E[\Delta Y_t^{11} \Delta Y_t^{11} > \Delta y_{t,prar}^{11}] - E[\Delta Y_t^{00} \Delta Y_t^{00} \leq \Delta y_{t,prc}^{00}]$
+ 3	$E[\Delta Y_t^{11}] - \frac{p_{at}}{p_c} \{E[\Delta Y_t^{11}] - E[\Delta Y_t^{01}]\} - E[\Delta Y_t^{00} \Delta Y_t^{00} > \Delta y_{t,prnt}^{00}]$	$E[\Delta Y_t^{11}] - \frac{p_{at}}{p_c} \{E[\Delta Y_t^{11}] - E[\Delta Y_t^{01}]\} - E[\Delta Y_t^{00} \Delta Y_t^{00} \leq \Delta y_{t,prc}^{00}]$
+ 4	$\frac{ATE_T}{p_c}$	$E[\Delta Y_t^{11}] - \frac{p_{at}}{p_c} \{E[\Delta Y_t^{11}] - E[\Delta Y_t^{01}]\} - E[\Delta Y_t^{00} \Delta Y_t^{00} \leq \Delta y_{t,prc}^{00}]$
Never-takers: $S_t = nt$		
1.a, 1.b, 2	$E[\Delta Y_t^{10}] - E[\Delta Y_t^{00} \Delta Y_t^{00} > \Delta y_{t,prc}^{00}]$	$E[\Delta Y_t^{10}] - E[\Delta Y_t^{00} \Delta Y_t^{00} \leq \Delta y_{t,prnt}^{00}]$
+ 3	$E[\Delta Y_t^{10}] - E[\Delta Y_t^{00} \Delta Y_t^{00} > \Delta y_{t,prc}^{00}]$	$E[\Delta Y_t^{10}] - E[\Delta Y_t^{00} \Delta Y_t^{00} \leq \Delta y_{t,prnt}^{00}]$
+ 4	$E[\Delta Y_t^{10}] - E[\Delta Y_t^{00} \Delta Y_t^{00} > \Delta y_{t,prc}^{00}]$	0

Note: ΔY_t^{zd} denotes first difference of outcomes in the observed group with $Z = z$ and $D_t = d$. $\Delta y_{t,pr_s}^{zd}$ denotes the pr_s -quantile of ΔY_t , where pr_s is equal to the relative size of stratum $S_t = s$ within the observed group with $Z = z$ and $D_t = d$.

the lower (upper) bound of the principal strata effect for never-takers, we thus assign the largest (or smallest) values of ΔY_t for individuals with $Z = 0$ and $D_t = 0$ to never-takers. Exactly the same approach can be used to bound the first-difference estimator for compliers, which we interpret as the income effect of the reform. Here, both components of Eq. (2) need to be bounded because compliers are observationally equivalent with never-takers in the observed group with $Z = 0$ and $D_t = 0$ and with always-takers in the observed group with $Z = 1$ and $D_t = 1$.

To further tighten these bounds, we can apply a set of additional assumptions that are predicted from our theoretical model:

Assumption 3. Exclusion restriction for always-takers

$$E[Y_t(0)|S_t = at] = E[Y_t(1)|S_t = at].$$

The exclusion restriction states that the reform has no effect on always-takers. This assumption is predicted by our theoretical model since the relevant part of the budget constraint for individuals who choose a partial DI benefit in the absence of the reform is the one to the right of the old notch. Neither the intercept nor the slope of this part of the budget constraint is affected by the reform. These individuals thus have no incentive to change labor supply at the intensive margin.⁶ This assumption is closely related to the standard IV approach (Angrist et al., 1996; Imbens and Angrist, 1994). The key difference is, however, that we apply the exclusion restriction only to always-takers and not to never-takers. Assumption 3 tightens the upper and lower bounds for the complier population since they are observationally equivalent to always-takers in the observed group with $Z = 1$ and $D_t = 1$.

Moreover, Assumptions 2 and 3 imply that the average effect of the reform is equal to the weighted average of the income and substitution effects:

$$E[Y_t(1) - Y_t(0)|Z = 1] = E[Y_t(1) - Y_t(0)|Z = 1, S_t = nt]Prob(S_t = nt|Z = 1) + E[Y_t(1) - Y_t(0)|Z = 1, S_t = c]Prob(S_t = c|Z = 1)$$

Assumption 4. Weak monotonicity of the mean potential outcomes within strata

$$E[\Delta Y_t|Z = 1, S_t = nt] \leq E[\Delta Y_t|Z = 0, S_t = nt]$$

$$E[\Delta Y_t|Z = 1, S_t = c] \geq E[\Delta Y_t|Z = 0, S_t = c]$$

⁶ The reform changed the intercept of the budget constraint, but it did so equally for individuals older and younger than fifty years. Formally it must hold that $U1' = U2'$, see Section 3.

The model predicts a negative substitution effect and a positive income effect. This assumption implies that the upper bound for the principal effect for never-takers is equal to zero (since our estimate for the average effect is positive, see Section 6.1), while the lower bound for the effect in the complier group becomes the standard Wald estimator.⁷ Table 2 summarizes the bounds, depending on the imposed assumptions.

5. Data and descriptive statistics

The analysis is based on the administrative data of the full sample of DI beneficiaries in Switzerland. We observe employment, earnings, DI benefits, disability degree, and background characteristics (earnings during contribution time, age, type of disability, canton of residence, marital status, citizenship, gender) in December of each year. The empirical analysis follows the stock of DI beneficiaries of the year 2003 up to 2007 for some sensitivity checks from 2001 to 2007. Individuals having congenital disorders are excluded since special rules apply for determining their disability degree and benefit level. The focus of this paper is on individuals who had a disability degree between 67 and 69% in 2003. This group represents only a small proportion of the full sample (3.4%).

In the main empirical analysis, we restrict the sample to individuals in an age bandwidth of 8 years around the threshold of 50, that is, aged 42 to 57 years in January 2004. Table 3 presents descriptive statistics for DI beneficiaries in the treatment and control groups for the year 2003. As the treatment group contains younger individuals, it is not surprising that it is characterized by higher earnings and a higher probability to work. While the disability degree is by construction very similar for the two groups, the treatment group is characterized by slightly lower earnings before disability onset and a higher probability to have a mental illness or an accident as a reason for DI, as opposed to musculoskeletal disease. Other background characteristics are balanced between the two groups.

The main outcome variables are employment (equal to one if the person has any earnings during the year) and yearly earnings, including zeros. While employment captures the extensive margin of the labor supply response, earnings captures the intensive and extensive margin.⁸ The trends in these outcomes and in the disability degree for the years 2001 to 2007 are presented in Fig. 2. Earnings (panel a) and labor supply (panel b) are higher for treated individuals than for

⁷ In Appendix A.2, we also discuss the general case of how this assumption affects bounds if the ATET is negative.

⁸ Note that one could analyze the intensive margin by focusing on positive earnings only. Our analysis is based on individual differences in earnings. We would thus lose any individuals who decide to enter the labor force because of the reform. We propose instead to divide the sample for a sensitivity check into those individuals that had some positive earnings in 2003 and those that did not. The results are presented in Table A.2 in Appendix A.3.

Table 3
Descriptive statistics for December 2003.

	Treated (1)	Control (2)
Panel A: Outcomes		
Annual earnings (CHF)	5265 (10,422)	4252 (9137)
Employment	0.35 (0.48)	0.29 (0.45)
Disability degree	67.8 (0.81)	67.8 (0.80)
Total DI benefits (CHF)	29,652 (11,884)	26,008 (9708)
Panel B: Independent variables		
Avg. earnings during contribution time	47,422 (26,787)	49,600 (23,949)
Mental illness	0.30 (0.46)	0.23 (0.42)
Musculoskeletal disease	0.29 (0.45)	0.39 (0.49)
Accident	0.20 (0.40)	0.15 (0.36)
Age	45.9 (2.27)	53.9 (2.26)
Married	0.64 (0.48)	0.67 (0.47)
Foreigner	0.33 (0.47)	0.32 (0.46)
Female	0.48 (0.50)	0.49 (0.50)
Observations	1336	2258

Note: Standard deviations are shown in parentheses. Earnings are equal to zero if the individual is not employed. Employment is a dummy variable equal to 1 if earnings are greater than zero. Disability degree denotes earnings loss with disability as a percentage of potential earnings without disability. Total DI benefits include main DI benefits, child benefits, spouse benefits, means-tested benefits and helpless allowance. Information on type of disability is available only for a subset of all observations (1335 treated and 2252 controls).

controls, but these variables follow a parallel development before the reform came into effect in 2004. After the reform, the gap between treated and control widens slightly (for better visibility, see Fig. A.3 in Appendix A.3, which present the time trends of normalized outcomes). The disability degree (panel c) is, by construction, equal in 2003 but increases strongly after the reform for the treated group while following a smooth time trend for the control group.

6. Results

6.1. Average effect of the reform

Estimates for the average effect of the reform for the year 2004 are presented in Table 4. Columns (1) and (2) show difference-in-differences estimates for the effect on earnings without and with the inclusion of background characteristics, respectively. Columns (3) and (4) show the estimates for the effect on employment, and columns (5) and (6) show the estimates for the effect on disability degree. The dependent variable is the change in the respective variable between the years 2004 and 2003. The coefficient on the treatment variable is therefore the difference-in-differences estimate.

Column (1) shows that average earnings decreased over time in both groups, which can be understood as an age effect or the effect of deteriorating health. The control group (represented by the constant) had decreased earnings of 453 CHF. The average effect of the reform on earnings is positive but very small and not statistically or economically significantly different from zero. Note, however, that the standard error of the coefficient is large, and the 95%-confidence

interval includes effects from –313 to +371 CHF, or an approximately 7% decrease or increase of annual earnings. This earnings effect is a mixture of the intensive and extensive margin responses. Table A.2 in the Appendix shows that the pure intensive margin earnings effect amounts to –540 CHF but does not remain statistically insignificant.

Columns (3) and (4) focus on the extensive margin. The causal effect of the reform amounts to a 2.3 percentage point increase in employment, which is not negligible given the baseline employment share of 35% in the treated population. This increase translates to a non-participation elasticity of roughly 0.26. Other studies that have investigated the effects of DI reforms estimated comparable elasticities between 0.12 and 0.36 (e.g. Gruber, 2000; Kostol and Mogstad, 2014).

Despite the presence of positive employment effects, columns (5) and (6) show that the reform leads to a large increase in the average disability degree. While the disability degree for the control group increases by 0.9 due to health deterioration, the disability degree of the treated group increases by an additional 2.9 percentage points, which is sufficient to reach the new threshold level of a full DI benefit.

One might argue that labor supply effects need longer time horizons to materialize, whereas disability degrees can be adjusted immediately in the year of the reform. Table A.3 presents the total effect of the reform for the years 2005 to 2007. The dependent variables are the differences with respect to the base year 2003 ($\Delta Y_{it} = Y_{it} - Y_{i0}$). The estimates for all years are comparable to the immediate impact. We find only modest, and in most cases insignificant, impacts on employment or earnings, but very stable and persistent effects on the disability degree and benefits.

6.2. Income and substitution effects

We now apply the bounds derived in Section 4.2 to estimate the income and substitution effects. Table 5 shows the respective strata proportions in column (1). Since the partial DI benefit indicator is zero for all individuals in the sample before the reform, this is equivalent to the cross-sectional estimator, where a binary indicator for partial DI benefit receipt after the reform is used as the dependent variable. The share of always-takers (represented by the constant) is very low at 0.6%. This means that in the absence of the reform, almost nobody increases their labor supply above the earnings threshold and receives a partial DI benefit.⁹ The reform increases the share of individuals who receive a partial DI benefit by 24.6 percentage points. These individuals are thus compliers to the reform. The large remaining share of individuals (74.8%) are never-takers.

Table 6 shows the estimated bounds for the income and substitution effects. The bounds imposing only parallel trends, independence of strata across treatment groups and monotonicity are wide and include zero. The estimated treatment effect on yearly earnings for compliers ranges from –2864 to 3730 CHF. The bounds for never-takers are tighter, ranging from –1001 to 861 CHF. Assuming the exclusion restriction for always-takers allows the bounds for compliers to be tightened only slightly to –2513 to 3135 CHF, leaving the bounds for never-takers unaffected. Assuming weak monotonicity of mean potential outcomes within strata (Assumption 4) implies that the upper bound for the principal effect for never-takers must be equal to zero since the average effect is slightly positive and that the lower bound for compliers is the standard Wald estimator. This result leads to informative bounds for the compliers that range from 136 to 3135 CHF.

⁹ These results implicitly confirm the findings from Bütler et al. (2015), who study the effect of a conditional cash program that is paid out to DI beneficiaries if they increase employment and lower DI benefits by at least one-quarter. The very low share of always-taker in our data is quantitatively very similar to the low take-up rate of the conditional cash payment of 0.5% documented by Bütler et al. (2015).

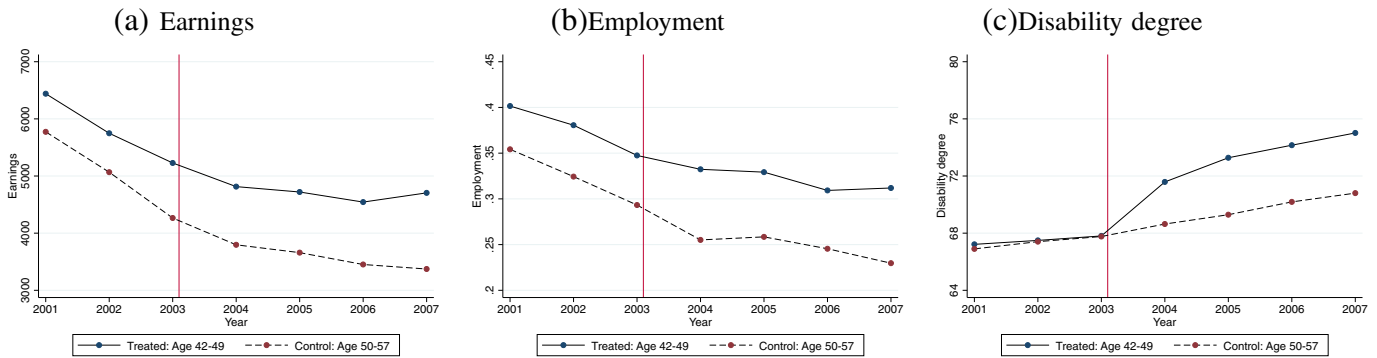


Fig. 2. Time trends for treated individuals and controls. Note: Solid lines represent the yearly mean outcomes for the treated group, and dashed lines represent control individuals. The sample consists of individuals aged between 42 and 57 years in January 2004 with a disability degree between 67 and 69% in December 2003. Outcomes are measured in December of each year. Vertical lines represent the time of the reform (January 2004).

Table 4
Total effect of the reform.

	Earnings		Employment		Disability degree	
	No controls	Controls	No controls	Controls	No controls	Controls
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	33.5 (173.2)	29.0 (174.6)	0.023** (0.009)	0.023** (0.009)	2.910*** (0.275)	2.887*** (0.271)
Constant	-452.7*** (102.2)	-24.5 (482.4)	-0.036*** (0.005)	-0.031 (0.025)	0.874*** (0.103)	1.821 (1.403)
R-squared	0.000	0.011	0.002	0.009	0.040	0.069
Observations	3581	3581	3581	3581	3581	3581

Note: Robust standard errors are shown in parantheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The sample consist of individuals aged between 42 and 57 years in January 2004 with a disability degree between 67 and 69% in December 2003. The outcomes are the first differences between the years 2004 and 2003. Earnings are the yearly earnings in CHF. Earnings are set to zero if not working. Employment is a dummy variable equal to 1 if the individual has earnings greater than zero. Disability degree is the potential earnings loss as a percentage of potential earnings without disability. The controls include dummies for canton of residence and year of observation, marital status, dummy for Swiss citizenship and gender.

The bounds for the employment effects for compliers are informative but very large. The reform has a positive impact on labor market participation for compliers of between 9.3 and 20.8 percentage points. This is a substantial increase compared to mean pre-reform labor market participation of 35% and amounts to an elasticity of between 1 and 1.7. To put this value into perspective, Marie and Vall Castello (2012) identify income effects in Spain and find an elasticity of labor force participation with respect to DI benefits of only 0.22. Note, however, that we identify the income effect for the subgroup of individuals where the income effect dominates, whereas Marie and Vall Castello (2012) estimate the average income effect for all DI applicants. For never-takers, the bounds cover the zero and

are much smaller. Never-takers reduce labor supply on the extensive margin by at most 3.7 percentage points. Interestingly, comparison of the employment and earnings effects can yield insight into the extensive and intensive margin responses of never-takers. The employment effect for never-takers is between -0.037 and 0 . Evaluated at the mean earnings of never-takers before the reform, the extensive margin effect can rationalize an earnings effect maximum of $-0.037 \cdot 5058 = -187$ CHF. However, we find a lower bound on the earnings effect of -1001 CHF. Therefore, approximately 19% of the earnings effect on never-takers can be explained by extensive margin responses. The remaining 81% of the earnings effect is driven by intensive margin responses.

We also predict bounds for the principal effects on the disability degree. Since the disability degree denotes the expected earnings loss due to a disability in relative terms, we expect a reverse sign for principal effects.¹⁰ We find tight bounds for never-takers, suggesting that the disability degree increases by 4 to 5 percentage points. For compliers, the bounds for the principal effects are somewhat larger, ranging from -1 to -5 percentage points. This difference reflects the fact that never-takers needed to increase their disability degree to keep a full DI benefit, whereas compliers are not forced to decrease their disability degree.

The timing of labor market responses can vary within different strata. For never-takers, we observe an immediate drop in disability

Table 5
Estimation of strata proportions.

	Partial DI benefit	
	No controls	Controls
	(1)	(2)
Treated	0.246*** (0.012)	0.249*** (0.012)
Constant	0.006*** (0.002)	-0.059 (0.039)
R-squared	0.161	0.184
Observations	3581	3581

Note: Robust standard errors are shown in parantheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The sample consist of individuals aged between 42 and 57 years in January 2004 with a disability degree between 67 and 69% in December 2003. The outcome is a dummy for partial DI benefit receipt in 2004. The controls include dummies for canton of residence and year of observation, marital status, dummy for Swiss citizenship and gender.

¹⁰ When deriving bounds for the effect on the disability degree, we need to take into account that always-takers decrease their disability degrees below 66% to receive a partial DI benefit, which affects the bounds that are predicted when considering Assumption 3 (exclusion restriction for always-takers).

Table 6
Bounds for the income and substitution effects.

Ass.	Earnings		Employment		Disability degree	
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
	(1)	(2)	(3)	(4)	(5)	(6)
Compliers: $S_t = c$						
1.a. 1.b. 2	-2864*** (427)	3730*** (513)	-0.090*** (0.023)	0.230*** (0.031)	-11.58*** (0.53)	-0.89*** (0.22)
+ 3	-2513*** (428)	3135*** (506)	-0.060*** (0.022)	0.208*** (0.030)	-4.86*** (0.53)	-1.07*** (0.22)
+ 4	136 (500)	3135*** (506)	0.093*** (0.030)	0.208*** (0.030)	-4.86*** (0.53)	-1.07*** (0.22)
Never-takers: $S_t = nt$						
1.a. 1.b. 2	-1001*** (207)	861*** (216)	-0.037*** (0.010)	0.052*** (0.012)	4.17*** (0.39)	5.42*** (0.36)
+ 3	-1001*** (207)	861*** (216)	-0.037*** (0.010)	0.052*** (0.012)	4.17*** (0.39)	5.42*** (0.36)
+ 4	-1001*** (207)	0	-0.037*** (0.010)	0	4.17*** (0.39)	5.42*** (0.36)

Note: Bootstrapped standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The sample consists of individuals aged between 42 and 57 years in January 2004 with a disability degree between 67 and 69% in December 2003. The outcomes are the first differences between the years 2004 and 2003. Employment is a dummy variable equal to 1 if the individual has earnings greater than zero. Earnings are the yearly earnings in CHF. Earnings are set to zero if not working. Disability degree is the potential earnings loss as a percentage of potential earnings without disability. The calculation of bounds for earnings and employment are as shown in Table 2. The calculation of bounds for disability degree take into account the fact that always-takers need to have a disability degree below 67%.

degree. To fulfill the necessary condition of having earnings below 30% of potential earnings without disability, we would also expect an immediate labor market reaction. This reaction is not observed: labor market responses, if they existed, were modest in the short and long term, suggesting that the necessary condition was already fulfilled and that individuals did not “bunch” incomes at 33% prior to the reform but already had lower incomes. In fact, we observe a large share of individuals in our sample with zero incomes. Compliers, on the other hand, experienced an immediate drop in DI benefits, and there could well be a time lag until they react by adjusting their labor market outcomes. Table A.4 in the Appendix shows the bounds analysis for the years 2005 to 2007. As expected, compliers respond somewhat stronger to the reform in the long term than in the short term. The bounds on income for the year 2007 show a minimum reaction of at least 1251 CHF up to a maximum of 5609 CHF, which is 80% more than the immediate reaction in 2004. By contrast, the bounds on employment remain almost constant, leading to the conclusion that compliers experienced wage increases over time rather than an increasing number of individuals becoming employed.

6.3. Discussion of the parallel trend assumption

The difference-in-differences identification strategy relies on the parallel-trend assumption, that is, the assumption that outcomes would develop in parallel for the control and treatment groups in the absence of the reform. This Section provides evidence for the identifying assumption and presents the robustness and sensitivity checks. For the first descriptive evidence, Fig. A.3 in the Appendix shows the time trends of earnings, employment, and disability degree with outcomes normalized to zero in the year 2003. The time trends are remarkably parallel for employment and disability degree and show only a slight deviation for earnings.

In the current setting, there are three main threats to the parallel-trend assumption. First, individuals could have potentially anticipated the introduction of the three-quarter DI benefit. Even though the political process was quick, and the three-quarter DI benefit was discussed only in the year 2002, anticipation effects may have led to a situation where individuals self-select in or out of the sample based on their age. For example, individuals who were younger than 50 years may have anticipated the reform and selected themselves

out of the sample by adjusting their labor supply accordingly. Fig. A.2 in Appendix A.3 shows that this is not a major issue. No discontinuity is observed in the age distribution among individuals with disability degrees between 67% and 69% in 2003. Moreover, Table A.6 shows the estimations when we specify our sample as the stock of insured individuals in the year 2001 instead of 2003, that is, before the reform was first discussed. Outcomes are then defined as the difference between December 2004 and December 2001. The estimated effects are slightly less precise but comparable in size to our baseline results. Finally, Table 7, Panel A provides evidence against anticipation by estimating the effects of a placebo reform implemented in 2003. None of the estimated effects are statistically significantly different from zero, and all the effects are close to zero in magnitude.

The second threat is that different age groups could experience different trends in labor market outcomes, even in the absence of any reform. One way to test if the estimated effects are driven by age differences between treated and control groups is to look at the effect of the reform on a sample with equal age differences, but for which the reform should have no effects. Table 7, Panel B shows the estimated

Table 7
Sensitivity: Placebo tests.

	Earnings	Employment	Disability degree
	(1)	(2)	(3)
Panel A: Placebo reform in 2003			
Treated	77.2 (266.6)	-0.009 (0.013)	0.163 (0.204)
R-squared	0.015	0.011	0.032
Observations	3089	3089	3089
Panel B: Disability degree between 70 and 75			
Treated	132.2 (97.8)	0.005 (0.013)	0.062 (0.053)
R-squared	0.005	0.001	0.004
Observations	13,642	13,642	13,642

Note: Robust standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The sample consist of individuals aged between 42 and 57 years in January 2003 with a disability degree between 67 and 69% in December 2002. The outcomes are first differences between the years 2003 and 2002.

effects of the reform on individuals with a disability degree between 70 and 75% in 2003. All workers, independent of their age, were subjected to the same features of the reform, namely, the threshold change for receiving a full DI benefit. However, no individual experienced a benefit loss because of the reform. The estimates do not reveal any significant effect.

Another way to determine whether the effects are driven solely by differences in age groups is to reduce the age bandwidth. We do this by employing a regression discontinuity design in first differences (FD-RDD) that estimates the local average treatment effect (LATE) directly at the age threshold of fifty. The identifying assumption is then that treatment variation is locally randomized for individuals close to the age threshold. This assumption implies that mean potential outcomes without treatment are identical for the treated and control groups

$$\lim_{\epsilon \rightarrow 0} E(Y_i(0)|A = 50 - \epsilon) = \lim_{\epsilon \rightarrow 0} E(Y_i(0)|A = 50 + \epsilon),$$

where A denotes age on January 1, 2004. The FD-RDD estimator can be implemented by estimating the regression $\Delta Y_{it} = \alpha_0 + \beta_1 Z_i + \beta_2 (A_i - 50) + \beta_3 (A_i - 50) Z_i + u_i$, where β_1 measures the local effect of the reform at age 50 years. Table 8 shows the results of the first-differenced RDD estimation. Although the effects are, in general, estimated somewhat less precisely, they are comparable in magnitude. Furthermore, Table A.5 checks the sensitivity of the results to reducing or increasing the age bandwidth in the DiD and FD-RDD regression and the sensitivity to a different functional form in the FD-RDD specification. The results are robust, except for the fact that employment effects are estimated to be somewhat lower and not statistically significantly different from zero in the FD-RDD regressions. While estimating the bounds in an FD-RDD setting is infeasible, Table A.8 shows that the results are insensitive to bandwidth changes.

Finally, in relation to the second threat, the parallel trends could be violated by the affect of age on the principal strata. To alleviate this concern, Table A.7 shows the strata proportions estimated by FD-RDD. The results for the strata proportions are comparable, with a slightly lower proportion of compliers. Based on this evidence, we believe that the parallel trend assumption holds in the present setting, for the estimation of the average effect and also for the income and substitution effects.

7. Why was the substitution effect so low?

Our empirical results partly support the predictions of the standard labor market model applied to DI: income effects play a significant role in explaining the high dependence on social assistance

and low labor supply among the disabled. However, we find only relatively small substitution effects, despite the fact that many individuals kept a full DI benefit. The key question is thus: why was the substitution effect of the reform so low? And how could individuals manage to increase their disability degree to keep a full pension without adjusting their labor market supply?

The simple labor market model presented in Section 3 relies on a set of assumptions. Most importantly, the model assumes that the only information used by the DI caseworker to set the disability degree is current earnings in relation to earnings before the onset of disability. Individuals can thus cheat the system by working less and consequently being classified as more disabled. This may not be a realistic assumption given that caseworkers have discretionary power to fix potential income with disability above the actual incomes. As such, our model provides the necessary (but not sufficient) condition to keep a full DI benefit. The necessary condition is that actual income must be strictly below the new threshold. The corresponding sufficient condition is that the caseworker uses his discretionary power and increases the disability degree.

Caseworkers used their discretionary power when fixing the disability degree even before the reform. In many cases, individuals have no earnings but a disability degree of less than 100%. For these individuals, the necessary condition for a higher disability degree was already fulfilled at the time of the reform, and no further labor market response was necessary. Therefore, caseworkers may have fixed the disability degree over the threshold of 70% to ensure that individuals kept full DI benefits after the reform, even in the absence of a change in labor supply. This is likely the case if caseworkers have incentives to increase the disability degree. A parliamentary report suggested “customer-centric” behavior of DI caseworkers, arguing that caseworkers have incentives to increase the disability degree above the next higher threshold to reduce workload – even though they were not advised to do so – especially since the risk of appeal was particularly high at the time of the implementation of the reform (FSIO, 2003; Janett et al., 2005).

Such “customer-centric” behavior requires a formal reassessment of the disability degree. The reform implemented a mandatory reassessment of disability degree for all individuals in our treatment group (FSIO, 2003). Fixing higher disability degrees was thus not overly costly to caseworkers but was part of a standard process. Therefore, the reform itself provided the possibility to adjust the disability degree. Unfortunately, we have no access to individual dossiers, so the interaction between the insured and the caseworkers remains unclear. Nevertheless, the available data provide some insights that this reassessment resulted in disability degree adjustments that were unrelated to any labor market response. Table 9, Panel A splits the sample into two groups to estimate the average effect of the reform. Those that earned 30% or less of pre-disability

Table 8
Sensitivity: FD-RDD.

	Earnings		Employment		Disability degree	
	No controls	Controls	No controls	Controls	No controls	Controls
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	134.0 (313.5)	161.6 (319.3)	0.014 (0.018)	0.016 (0.018)	3.180*** (0.493)	3.162*** (0.488)
Constant	-440.1* (232.4)	-21.5 (536.8)	-0.032** (0.013)	-0.028 (0.027)	0.727*** (0.238)	1.762 (1.458)
R-squared	0.000	0.011	0.002	0.009	0.041	0.070
Observations	3581	3581	3581	3581	3581	3581

Note: Robust standard errors are shown in parantheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The sample consist of individuals aged between 42 and 57 years in January 2004 with a disability degree between 67 and 69% in December 2003. The outcomes are first differences between the years 2004 and 2003. Earnings are the yearly earnings in CHF. Earnings are set to zero if not working. Employment is a dummy variable equal to 1 if the individual has earnings greater than zero. Disability degree is the potential earnings loss as a percentage of potential earnings without disability. The controls include dummies for canton of residence and year of observation, marital status, dummy for Swiss citizenship and gender.

Table 9
Effects of the reform on the disability degree.

Panel A: Necessary condition			
Dependent: Disability degree	Original sample	Non-binding necessary condition	Binding necessary condition
Treated	2.887*** (0.271)	3.211*** (0.302)	0.774*** (0.526)
Observations	3581	3122	459
R-squared	0.070	0.082	0.097
Panel B: Targeting			
Dependent:	Increase in dd	Increase in dd no decrease in earnings	Increase in dd no decrease in earnings exactly to cash cliff
Treated	0.246*** (0.013)	0.189*** (0.013)	0.053*** (0.007)
Observations	3581	3581	3581
R-squared	0.140	0.102	0.037

Note: Robust standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The sample consists of individuals aged between 42 and 57 years in January 2004 with a disability degree between 67 and 69% in December 2003. Panel A: Individuals with a non-binding necessary condition have an analytical disability degree of 70% or higher in December 2003. Individuals with a binding necessary condition have an analytical disability degree of 69% or lower in December 2003. The analytical disability degree is calculated as current earnings divided by earnings during the contribution period. Panel B: Outcomes are binary indicator variables capturing (1) an increase in disability degree; (2) an increase in disability degree without a corresponding decrease in earnings; and (3) an increase in disability degree to exactly 70% between the years 2004 and 2003. Disability degree is the potential earnings loss as a percentage of potential earnings without disability. Controls include dummies for canton of residence, year of observation, marital status, Swiss citizenship and gender.

earnings before the reform, and thus are not constrained by the necessary condition, and those that earned 31% or more and for whom the necessary condition is binding. Pre-disability earnings are an imperfect proxy for the earnings potential without disability, and we can therefore argue that for those who earned 31% or more, an earnings decrease is necessary to fulfill the necessary condition after the reform. The observed pattern is insightful. Only the group of individuals who already fulfilled the necessary condition before the reform significantly increased their disability degree by 3.2 points on average. Individuals who would have needed to adjust their labor market behavior first to fulfill the necessary condition did not increase their disability degrees strongly. These results suggest that at least a portion of the disability degree increases was driven by caseworker reactions exploiting their leeway in setting disability degrees and less by the actual labor market reactions of disabled individuals.

We further investigate potential caseworker reactions in Table 9, Panel B. Column (1) runs a regression on the binary outcome of an increase in the disability degree between 2003 and 2004. Individuals in the treatment group have a 25 percentage point higher probability of increasing their disability degree than individuals in the control group. The shift in the disability degree thus falls into the time of the mandatory reassessment process implemented by the DI system. Column (2) focuses on those disability degree increases that occur without any corresponding decrease in earnings. While only 5.6% of all non-treated individuals experience such an increase, 24.5% of all treated individuals increase their disability degree without lowering earnings. The majority of individuals in our treatment group (approximately 80%) whose disability degree was increased had no corresponding decrease in earnings. Finally, column (3) investigates how many individuals increase their disability degree by exactly the amount necessary to reach the new cash cliff at 70. A high share (approximately 30%) of those who increased their disability degree without a corresponding decrease in earnings increased their disability degree by only the small amount required to reach the next threshold level. All these estimated effects are highly statistically significant and are not related to labor supply changes but purely to the usage of the discretionary power a caseworker has when fixing the disability degree.

Taken together, this additional evidence is consistent with “customer-centric” behavior of DI caseworkers, which likely reduced the welfare burden of the reform for those DI beneficiaries that were affected by the benefit cut. When designing future reforms of the DI system, such behavior should be considered.

8. Conclusion

This paper evaluates a reform of the Swiss DI system that introduced a three-quarter DI benefit and thus further graduated the existing system. The analysis focuses on those DI beneficiaries who, based on their pre-reform disability degree, would lose a one-quarter DI benefit and are therefore faced with a lower earnings threshold to remain eligible for the full DI benefit.

The main takeaways of the paper are twofold. First, the average labor supply effects on the intensive and extensive margin are modest and in line with existing literature. The average effect consists of conflicting income and substitution effects. We partially identify these effects. The bounds for the income effects suggest that individuals who complied with the reform (approximately 25%) increased yearly earnings by between 136 and 3135 CHF. This is a sizable amount compared to previous earnings (max 50% increase in earnings). The bounds for the substitution effect indicate that individuals who kept a full DI benefit reduced yearly earnings by a maximum of 1000 CHF (if any). There is a huge public interest in reforms that remove existing work disincentives caused by cash cliffs. A widely discussed reform is to further graduate the DI benefit payout structure. Policymakers need to be aware that this policy can lead to conflicting income and substitution effects, making the average effect ambiguous. In summary, such policies should be implemented only if the substitution effect is severe and many DI beneficiaries are cash-cliff constrained. Our paper shows that the substitution effect is not the driving force imposing work disincentives on DI beneficiaries – at least in Switzerland in the current setting with strong discretionary power of the caseworkers. Part of this conclusion is driven by the fact that many individuals with a disability degree between 67 and 69% do not work at all, even before the reform. This situation, however, might be a direct consequence of the relatively large

income effect found in this paper. The replacement rates guaranteed by different non-earned income sources are very high in Switzerland. The income effect may therefore be the driving force leading to low labor supply and high dependency on DI benefits among the disabled.

The second main result of the paper is that there is a large effect of the reform on the average disability degree. Our analysis shows that this effect is driven by individuals who already fulfilled the necessary condition of a higher disability degree before the reform, that is, individuals who earned less than 30% of their potential earnings without disability and oftentimes were not employed at all. The fact that a large proportion of the disability degree increases were exactly the amount that allowed individuals to keep a full DI benefit suggests that caseworkers act in customer-centric manner. This was eased in the specific context of the reform by an automatic reevaluation of the disability degree of all individuals in the treatment group, even though caseworkers were not advised to increase disability degrees (FSIO, 2003; Janett et al., 2005). Linking the payout structure directly to disability induced income losses – together with the considerable leeway of caseworkers when assessing potential incomes with and without disability – is thus a doubtful concept that will lead to manipulation of the disability degree and undermine the desired incentive effects of a graduated pension scheme.

We should, however, not overgeneralize the results to the potential immediate effects of the introduction of a partial DI benefit schedule. Employers in the Swiss labor market are well aware of the discontinuities of the budget constraint and may be willing to help strategic bunching by adjusting work contracts accordingly. This may not be the case with a new introduction of a partial DI benefit schedule. Therefore, more time may be required until individuals sort optimally.

This paper provides a framework how to elicit income and substitution effects in a partial DI benefit system. However, the analysis is performed in a setting with strong caseworker discretionary power. A valuable direction for future research would thus be to elicit the income and substitution effects in different settings where the individual reactions of the DI beneficiaries can be clearly distinguished from potential caseworker reactions.

Appendix A. Appendix

A.1. Point identification of strata proportions

Principal strata and the mix of principal strata observed in groups with values $P_t = p$ and $D = d$ are presented in Table A.1. **Assumption 2 (monotonicity)** rules out the existence of defiers. We can therefore directly observe the population proportion of never-takers and always-takers in the treated and control groups:

$$p_{nt|D=1} = Pr(P_t = 0|D = 1)$$

$$p_{at|D=0} = Pr(P_t = 1|D = 0)$$

Assumption 1.b (Independence of potential DI benefit receipt) states that the potential DI benefit level, and thus principal strata, is independent of exposure to the reform:

$$p_{nt} = p_{nt|D=1} = p_{nt|D=0}$$

$$p_{at} = p_{at|D=1} = p_{at|D=0}$$

This assumption enables the point identification of the population proportion of compliers:

$$p_c = Pr(P_t = 1|D = 1) - Pr(P_t = 1|D = 0)$$

Table A.1

Types of individuals in potential treatment and observed treatment groups.

Potential DI benefit level			Observed groups		
$P_t(1)$	$P_t(0)$		P_t	D	
0	0	Never-taker	1	1	Complier or Always-taker
1	0	Complier	0	0	Complier or Never-taker
0	1	Defier	0	1	Defier or Never-taker
1	1	Always-taker	1	0	Defier or Always-taker

A.2. Partial identification of principal effects

In this paper, we seek to estimate the causal effect of exposure to the reform on labor supply in different latent groups. **Assumption 1.a (parallel trends within strata)** enables the effect of interest to be rewritten as the first-difference estimator in stratified samples:

$$E[Y(1)-Y(0)|D = 1, S_t = s] = E[\Delta Y_t|D = 1, S_t = s] - E[\Delta Y_t|D = 0, S_t = s]$$

Under **Assumption 2 (monotonicity)**, we can directly observe the first component of the first-difference estimator for compliers ($E[\Delta Y_t|D = 1, S_t = nt] = E[\Delta Y_t|D = 1, P_t = 0] = E[\Delta Y_t^{10}]$) and the second component for always-takers ($E[\Delta Y_t|D = 0, S_t = at] = E[\Delta Y_t|D = 0, P_t = 1] = E[\Delta Y_t^{01}]$). The remaining components of the first-difference estimators cannot be observed because compliers are observationally equivalent with always-takers in the group with $P_t = 1$ and $D = 1$ and with never-takers in the group with $P_t = 0$ and $D = 0$.

We construct worst-case scenarios for the remaining unobserved components by recognizing that the average ΔY_t for individuals in mixed observed groups (such as $P_t = 1$ and $D = 1$) can be written as

$$E[\Delta Y_t^{11}] = \frac{p_{at}}{p_{at} + p_c} E[\Delta Y_t|D = 1, S_t = at] + \frac{p_c}{p_{at} + p_c} E[\Delta Y_t|D = 1, S_t = c]$$

Since strata proportions are point identified, $E[\Delta Y_t|D = 1, S_t = at]$ can be bounded from above (below) by the $pr_{at} = \frac{p_{at}}{p_{at} + p_c}$ fraction of the largest (smallest) values of ΔY_t for individuals in the observed group with $P_t = 1$ and $D = 1$. The resulting worst-case bounds for always-taker are then equal to

$$LB_{at}^{wc} = E[\Delta Y_t^{11} | \Delta Y_t^{11} \leq \Delta y_{t,pr_{at}}^{11}] - E[\Delta Y_t^{01}]$$

$$UB_{at}^{wc} = E[\Delta Y_t^{11} | \Delta Y_t^{11} > \Delta y_{t,pr_c}^{11}] - E[\Delta Y_t^{01}],$$

where $\Delta y_{t,pr_s}^{pd}$ denotes the pr_s quantile of ΔY_t in the group with $P_t = p$ and $D = d$.

We follow the same approach to bound the principal effect for never-takers:

$$LB_{nt}^{wc} = E[\Delta Y_t^{10}] - E[\Delta Y_t^{00} | \Delta Y_t^{00} > \Delta y_{t,pr_c}^{00}]$$

$$UB_{nt}^{wc} = E[\Delta Y_t^{10}] - E[\Delta Y_t^{00} | \Delta Y_t^{00} \leq \Delta y_{t,pr_{nt}}^{00}]$$

For compliers, we take into account the fact that both components of the first-difference estimator need to be bounded in the same manner:

$$LB_c^{wc} = E[\Delta Y_t^{11} | \Delta Y_t^{11} \leq \Delta y_{t,prc}^{11}] - E[\Delta Y_t^{00} | \Delta Y_t^{00} > \Delta y_{t,prnt}^{00}]$$

$$UB_c^{wc} = E[\Delta Y_t^{11} | \Delta Y_t^{11} > \Delta y_{t,prnt}^{11}] - E[\Delta Y_t^{00} | \Delta Y_t^{00} \leq \Delta y_{t,prc}^{00}]$$

Assumption 3 (Exclusion restriction) states that the principal effect for always-takers is zero, which point identifies the first component of the treatment effect for always-takers

$$E[\Delta Y_t | D = 1, S_t = at] = E[\Delta Y_t | D = 0, S_t = at] = E[\Delta Y_t^{01}]$$

This allows us to point identify the first component of the first-difference estimator for compliers via the observation that the observed ΔY_t in the group with $P_t = 1$ and $D = 1$ is the weighted average of ΔY_t for compliers and always-takers:

$$E[\Delta Y_t | D = 1, S_t = c] = E[\Delta Y_t^{11}] - \frac{p_{at}}{p_c} \{E[\Delta Y_t^{11}] - E[\Delta Y_t^{01}]\}$$

The bounds for compliers are thus tightened to

$$LB_c^{+A3} = E[\Delta Y_t^{11}] - \frac{p_{at}}{p_c} \{E[\Delta Y_t^{11}] - E[\Delta Y_t^{01}]\} - E[\Delta Y_t^{00} | \Delta Y_t^{00} > \Delta y_{t,prnt}^{00}]$$

$$UB_c^{+A3} = E[\Delta Y_t^{11}] - \frac{p_{at}}{p_c} \{E[\Delta Y_t^{11}] - E[\Delta Y_t^{01}]\} - E[\Delta Y_t^{00} | \Delta Y_t^{00} \leq \Delta y_{t,prc}^{00}]$$

Assumption 3 does not affect the bounds for never-takers because they never share an observed group with always-takers

$$LB_{nt}^{+A3} = LB_{nt}^{+wc}$$

$$UB_{nt}^{+A3} = UB_{nt}^{+wc}$$

Assumption 3 furthermore allows the average effect to be rewritten as a weighted average of the principal stratum effects for compliers and never-takers:

$$\begin{aligned} ATET_t &= E[Y_t(1) - Y_t(0) | D = 1] \\ &= E[Y_t(1) - Y_t(0) | D = 1, S_t = nt] \text{Prob}(S_t = nt) \\ &\quad + E[Y_t(1) - Y_t(0) | D = 1, S_t = c] \text{Prob}(S_t = c) \\ &= \{E[\Delta Y_t | D = 1, S_t = nt] - E[\Delta Y_t | D = 0, S_t = nt]\} p_{nt} \\ &\quad + \{E[\Delta Y_t | D = 1, S_t = c] - E[\Delta Y_t | D = 0, S_t = c]\} p_c \end{aligned}$$

Assumption 4 (Weak monotonicity) predicts the sign of the principal stratum effects, i.e., $E[Y(1) - Y(0) | D = 1, S_t = nt] \leq 0$ and $E[Y(1) - Y(0) | D = 1, S_t = c] \geq 0$. Therefore, the upper bound for never-takers and/or the lower bound for compliers is equal to zero, depending on the sign of the $ATET_t$. On the basis of the observation that the average effect is equal to the weighted average of the principal stratum effects, the corresponding upper (lower) bound for compliers (never-takers) is equivalent to the standard Wald estimator, i.e., the average effect of the reform in the full population weighted by the inverse of the probability that a person belongs to the respective stratum. In the case that $ATET_t = 0$, the lower bounds for compliers and the upper bound for never-takers are both zero.

The upper bounds for compliers and the lower bounds for never-takers, however, are not affected. For compliers, the bounds thus become:

$$LB_c^{+A4} = \max\left(0, \frac{ATET_t}{p_c}\right)$$

$$UB_c^{+A4} = UB_c^{+A3}$$

The bounds for never-takers are:

$$LB_{nt}^{+A4} = UB_{nt}^{+A3}$$

$$UB_{nt}^{+A4} = \min\left(0, \frac{ATET_t}{p_{nt}}\right)$$

Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpubeco.2018.12.001>.

References

- J.D., Angrist, G.W., Imbens, D.B., Rubin, 1996. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* 91 (434), 444–455.
- D.H., Autor, M.G., Duggan, 2007. Distinguishing income from substitution effects in disability insurance. *Am. Econ. Rev.* 97 (2), 119–124.
- D.H., Autor, M.G., Duggan, 2010. Supporting work: a proposal for modernizing the U.S. disability insurance system. Project report. The Center for American Progress and The Hamilton Project.
- L., Borghans, A.C., Gielen, E.F., Luttmer, 2014. Social support substitution and the earnings rebound: evidence from a regression discontinuity in disability insurance reform. *Am. Econ. J. Econ. Pol.* 6 (4), 34–70.
- J., Bound, 1989. The health and earnings of rejected disability insurance applicants. *Am. Econ. Rev.* 79 (3), 482–503.
- J., Bound, R.V., Burkhauser, 1999. Economic analysis of transfer programs targeted on people with disabilities. *Handb. Labor Econ.* 3, 3417–3528.
- K.M., Brown, 2013. The link between pensions and retirement timing: lessons from California teachers. *J. Public Econ.* 98, 1–14.
- M., Büttler, E., Deuchert, M., Lechner, S., Staubli, P., Thiemann, 2015. Financial work incentives for disability benefit recipients: lessons from a randomised field experiment. *IZA J. Labor Policy* 4 (1), 18.
- M., Campolieti, C., Riddell, 2012. Disability policy and the labor market: evidence from a natural experiment in Canada, 1998–2006. *J. Public Econ.* 96, 306–316.
- S., Chen, W., van der Klaauw, 2008. The work disincentive effects of the disability insurance program in the 1990s. *J. Econ.* 142 (2), 757–784.
- C.E., Frangakis, D.B., Rubin, 2002. Principal stratification in causal inference. *Biometrics* 58, 21–29.
- E., French, J., Song, 2014. The effect of disability insurance receipt on labor supply. *Am. Econ. J. Econ. Pol.* 6 (2), 291–337.
- FSIO, 2003. IV-Rundschreiben Nr. 183 vom 9. Oktober 2003. Bundesamt für Sozialversicherungen BSV.
- A., Gelber, T.J., Moore, A., Strand, 2017. The effect of disability insurance payments on beneficiaries' earnings. *Am. Econ. J. Econ. Pol.* 9 (3), 1–33.
- J., Gruber, 2000. Disability insurance benefits and labor supply. *J. Polit. Econ.* 108 (6), 1162–1183.
- G.W., Imbens, J.D., Angrist, 1994. Identification and estimation of local average treatment effects. *Econometrica* 62 (2), 467–475.
- D., Janett, S., Zogg, H., Heinis, 2005. Faktoren des Rentenwachstums in der Invalidenversicherung. Technical report. Parlamentarische Verwaltungskontrolle.
- A., Karlström, M., Palme, I., Svensson, 2008. The employment effect of stricter rules for eligibility for DI: evidence from a natural experiment in Sweden. *J. Public Econ.* 92 (10–11), 2071–2082.
- L., Kauer, 2014. The effect of cutting disability insurance benefits on labor supply in households. Discussion Paper Series. University of St. Gallen.
- P., Kline, M., Tartari, 2016. Bounding the labor supply responses to a randomized welfare experiment: a revealed preference approach. *Am. Econ. Rev.* 106 (4), 972–1014.
- A.R., Kostol, M., Mogstad, 2014. How financial incentives induce disability insurance recipients to return to work. *Am. Econ. Rev.* 104 (2), 624–655.
- Liebert, H., 2016. Medical screening and award errors in disability insurance, mimeo, University of St. Gallen.
- N., Maestas, K.J., Mullen, A., Strand, 2013. Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt. *Am. Econ. Rev.* 103 (5), 1797–1829.

- O., Marie, J., Vall Castello, 2012. Measuring the (income) effect of disability insurance generosity on labour market participation. *J. Public Econ.* 96 (1), 198–210.
- T.J., Moore, 2014. The employment effect of terminating disability benefits. NBER Working Papers. National Bureau of Economic Research, Inc.
- OECD, 2010. *Sickness, Disability and Work. Breaking the barriers. A Synthesis of Findings A OECD Countries.* OECD.
- Ruh, P., Staubli, S., 2015. Financial Incentives and Earnings of Disability Insurance Recipients: Evidence from a Notch Design, mimeo.
- E., Saez, 2010. Do taxpayers bunch at kink points? *Am. Econ. J. Econ. Pol.* 2, 180–212.
- J., Schimmel, D.C., Stapleton, J.G., Song, 2011. How common is “parking” among social security disability insurance beneficiaries? Evidence from the 1999 change in the earnings level of substantial gainful activity. *Soc. Secur. Bull.* 71 (4), 77–92.
- S., Staubli, 2011. The impact of stricter criteria for disability insurance on labor force participation. *J. Public Econ.* 95 (9–10), 1223–1235.
- T., von Wachter, J., Song, J., Manchester, 2011. Trends in employment and earnings of allowed and rejected applicants to the social security disability insurance program. *Am. Econ. Rev.* 101 (7), 3308–3329.
- R.R., Weathers, J., Hemmeter, 2011. The impact of changing financial work incentives on the earnings of social security disability insurance (SSDI) beneficiaries. *J. Policy Anal. Manag.* 30 (4), 708–730.