Empirical Economics
https://doi.org/10.1007/s00181-018-1587-4

Disability insurance benefits and labor supply decisions: evidence from a discontinuity in benefit awards

Tobias Müller[1,2] **· Stefan Boes**[3]

Received: 28 February 2018 / Accepted: 2 November 2018

Abstract
The effect of disability insurance (DI) benefits on the labor supply of individuals is a disputed topic in both academia and policy. We identify the impact of DI benefits on working full-time, working part-time or being out of the labor force by exploiting a discontinuity in the DI benefit award rate in Switzerland above the age of 56. Using rich survey data and a discrete endogenous switching model, we find that DI benefit receipt increases the probability of working part-time by approximately 32% points, decreases the probability of working full-time by approximately 35% points and has little effect on the probability of being out of the labor force for the average beneficiary. Looking at the treatment effect distribution, we find that male, middle- to high-income and relatively healthy DI beneficiaries are more likely to adjust their labor supply from full-time to part-time, whereas women, low-income and ill beneficiaries tend to drop out of the labor market. Our results shed new light on the mechanisms explaining low DI outflow rates and may help better target interventions.

Keywords Disability insurance · Labor market participation · Fuzzy regression discontinuity design · Discrete endogenous switching models · Maximum simulated likelihood

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00181-018-1587-4) contains supplementary material, which is available to authorized users.

✉ Stefan Boes
stefan.boes@unilu.ch

Tobias Müller
tobias.mueller@vwi.unibe.ch

[1] Department of Economics, University of Bern, Schanzeneckstrasse 1, PO Box 3001, Bern, Switzerland

[2] School of Management and Law, Zurich University of Applied Sciences (ZHAW), Bahnhofplatz 12, 8400 Winterthur, Switzerland

[3] Department of Health Sciences and Health Policy, Center for Health, Policy and Economics, University of Lucerne, Frohburgstrasse 3, PO Box 4466, 6002 Lucerne, Switzerland

Published online: 17 November 2018

JEL Classification J2 · C35 · C36

1 Introduction

In most industrialized countries, the costs of disability insurance (DI) are substantial and place a heavy burden on public finances. Many workers leave the labor market permanently due to health issues, driving both the inflow and stock of DI beneficiaries. In fact, the number of DI beneficiaries as a share of the working-age population (the disability recipiency rate) has risen rapidly over the past decades across the developed world: From 1970 to 2013, the average annual growth rate in disability recipiency was 3.1% in the USA, 2.1% in Great Britain, 3.0% in Australia and 2.7% in Sweden (Burkhauser et al. 2013). In Switzerland, the focus of this study, the number of beneficiaries has risen from approximately 199,000 in 2000 to 230,000 in 2013 (average annual growth rate of 1.1%). The trends in DI beneficiaries are reflected in DI expenditure levels. For instance, the DI cash transfer payments in the USA totaled $25 billion in 1990, rising to $140 billion in 2013. Similarly, DI expenditures in Switzerland increased from 4.1 billion Swiss Francs in 1990 to 9.3 billion Swiss Francs in 2013.

As a response to the unsustainable growth in DI costs and rising number of beneficiaries, policy makers have introduced various reforms addressing the incentive structure of the DI system. However, policies have focused mainly on reducing DI inflow by offering employment-creating measures, such as job placement and career advice for applicants (OECD 2010). At the same time, countries have tightened the access to benefits and reduced compensation generosity. Despite these efforts, little has been done to address the stock of existing beneficiaries, which is surprising given that DI outflow for reasons other than death is as low as 1% across the OECD (OECD 2010). The low outflow is suggestive of lock-in effects that keep beneficiaries out of the labor force even though some of them may have valuable work capacities. Economic theory suggests a moral hazard issue as a potential explanation for such lock-in effects: Individuals who receive a high disutility from working remain out of the labor force because the DI system redistributes resources to them (Bound and Burkhauser 1999; Shu 2015). The negative effects of DI benefits on labor market participation are well documented in the empirical literature (e.g., Bound 1989; Chen and Van der Klaauw 2008; Marie and Vall Castello 2012; French and Song 2014; Moore 2015; Frutos and Vall Castello 2015), although evidence for Switzerland is scarce (notable exceptions are Kauer 2014, and Eugster and Deuchert 2015).

In the light of the above, it is important to learn more about the mechanisms contributing to a low DI outflow and the incentives embodied in the DI system. In this paper, we investigate how existing beneficiaries in Switzerland adapt their labor supply decisions as a response to receiving DI benefits. We address the endogeneity of DI benefit status by exploiting a discontinuity in the benefit award rate. Applicants aged 56 or above are much more likely to receive DI benefits than people below that age. This discontinuity in the benefit award rate arises due to the common practice of DI offices of using the age of an applicant as a key factor when deciding upon DI benefits (Federal Social Insurance Office 2013a, b). The rationale here is that DI offices take into account that older workers have more problems reentering the job market once they encounter health issues than younger workers, and thus, DI is often used as a substitute for early retirement. Given that DI applicants cannot manipulate their age, the discontinuity in the benefit award rate can serve as an exogenous instrument for DI benefit receipt.

We estimate the impact of DI benefits on the decision to work full-time, part-time or be out of the labor force using a discrete endogenous switching (ES) model (Miranda and Rabe-Hesketh 2006; Roodman 2011). Under the assumption of jointly normal errors, the model allows us to estimate the distribution of treatment effects on the treated. In contrast, two-stage least squares and fuzzy regression discontinuity (RD) estimations only produce local average treatment effects and cannot readily be applied to multinomial outcomes (Angrist and Pischke 2009). A practical issue with discrete ES models is the rather complicated likelihood function, which must be maximized using simulated likelihood methods (MSL methods).[1] To the best of our knowledge, this paper is the first to examine the effects of DI benefits on a multinomial outcome of labor supply using a discrete ES model, which we see as an important contribution not only in the Swiss context but also more generally. The existing literature (e.g., Gruber 2000; Chen and Van der Klaauw 2008; Marie and Vall Castello 2012; French and Song 2014; Moore 2015; see also below) either investigates the effects of DI benefits on the extensive margin, i.e., on the binary decision to work versus not to work, neglecting transitions from full-time to part-time employment that are especially relevant for DI systems with partial benefits, like the one in Switzerland, or evaluates the effects at the intensive margin, i.e., the hours of work for those who actually participate in the labor market, which might suffer from selection issues (Angrist and Pischke 2009). Modeling the more complex multinomial working decision allows us to investigate both intensive and extensive margin effects and thus has the potential to reveal new facets of the incentives embodied in the DI system.

The results of our analysis reveal a strong impact of DI benefits on the labor market decisions of existing DI beneficiaries. In particular, we find that individual DI benefit receipt significantly increases the probability of working in a part-time job by 32% points, reduces the probability of working full-time by 35% points and has little effect on the probability of being out of the labor force for the average beneficiary. These findings suggest that DI benefits do not necessarily force beneficiaries out of the labor force but instead induce a shift in the labor supply from full-time to part-time work. We also find evidence for substantial effect heterogeneity. Comparing the characteristics of the existing beneficiaries in different quartiles of the treatment effect distribution, we find that men are more likely to shift their labor supply from full-time to part-time, whereas women tend to drop out of the labor force. Low-income and very ill individuals also tend to drop out of the labor force, whereas middle- to high-income and relatively healthy DI beneficiaries tend to shift their labor supply from working full-time to part-time. Thus, our estimations of the discrete ES model reveal new aspects of the heterogeneous effects of DI benefits at the intensive and extensive margins of labor supply that were not captured by the earlier literature. Our results help better explain

[1] This will be done using the Geweke, Hajivassiliou and Keane (GHK) algorithm that allows estimating higher-order dimensional cumulative normal distributions (Roodman 2011).

the low DI outflow in Switzerland, and possibly other OECD countries, and we provide novel evidence for policy makers that might prove useful in planning interventions that are targeted at particular subgroups of the population.

The paper proceeds as follows. In Sect. 2, we briefly review the different empirical strategies that have been used in the literature to identify the effects of DI incentives on labor force participation. In Sect. 3, we discuss the structure of the Swiss DI system and give information on the eligibility determination process relevant for our study. Section 4 describes the data and variables used for the analysis and sheds light on the population of beneficiaries. The empirical strategy and related validity checks, the discrete ES model and estimation procedure of our study are outlined in Sect. 5. Section 6 presents the results for the average beneficiary and discusses effect heterogeneity. Final conclusions are drawn in Sect. 7.

2 Review of the related literature

Different identification strategies have been used in the past to model individual responses to DI programs. Early empirical studies estimate labor force participation (LFP) equations using standard regression techniques. For example, Parsons (1980a, b) uses cross-sectional data from the National Longitudinal Survey of Older Men to estimate labor force non-participation as a function of the SSDI[2] replacement rate and characteristics such as age, gender, education and health status. The results suggest an elasticity of labor force non-participation with respect to benefit levels[3] for men aged 45–59 of between 0.49 (1980a) and 0.93 (1980b).

One of the most influential papers to this day is the study by Bound (1989). Bound suggests that SSDI applicants who fail to pass the medical screening necessary to qualify for the program form a natural control group for beneficiaries. As a rationale, he argues that rejected SSDI applicants and beneficiaries should be similar with respect to observed and unobserved characteristics, thus making them comparable in their LFP decisions. Using data from the 1972 Survey of Disabled and Non-Disabled Adults (SDNA) and the 1978 Survey of Disability and Work (SDW), his analysis shows that less than one-third of the rejected applicants were working at the time of the survey and less than 50% had worked at some point in the previous year. Bound argues that the rejected applicants are healthier and more capable of work than those who receive an income transfer, and thus, their LFP rate forms an upper bound for the work participation behavior of the beneficiaries in the absence of the disability benefits.

More recent studies use natural experiments to shed light on the LFP effects of DI benefits. For example, Gruber (2000) exploits a large policy change in the Canadian DI system in the late 1980s. DI in Canada operates basically the same as in the USA but with the key difference that there are two distinct DI programs: the Quebec Pension Plan (QPP), which covers only the region of Quebec, and the Canada Pension

[2] Social Security Disability Insurance (SSDI) is one of the two main federal programs that provide cash assistance to the disabled in the USA. The other program is Supplemental Security Income (SSI), which was introduced in 1972 to provide a minimum level of income to impaired individuals.

[3] The elasticity of non-participation with respect to benefits is defined as the ratio of the change in labor supply relative to the change in potential benefits (Gruber 2000).

Plan (CPP), which covers the remaining regions of Canada. Until 1986, the QPP was substantially more generous in terms of benefits than the CPP. In 1987, the CPP raised its benefits by 36% to the level of the QPP to equalize the generosity levels of the two systems. Gruber uses this exogenous policy change in disability benefits to estimate the labor supply effects using a difference-in-differences approach and finds an elasticity of non-participation with respect to benefit levels of between 0.28 and 0.36.

Staubli (2011) analyzes the impact of a tightening in disability eligibility rules on the labor supply of older workers in Austria. He uses the policy change introduced by the Structural Adjustment Act in 1996, which led to stricter disability eligibility criteria for men age 55–57, to examine (i) how tighter criteria for benefits affect enrollment and employment and (ii) whether a tightening in eligibility rules leads to spillover effects into other programs. Relying on a DID approach, he finds that the share of beneficiaries in the affected age group significantly decreased by 6–7.2% points after the reform was implemented. Additionally, his estimates indicate an increase in employment by 1.7–3.4% points after the policy change. At the same time, his results suggest an increase in the share of individuals receiving unemployment or disability insurance benefits indicating substantial spillover effects. Studies concerning the effects of screening stringency on LFP in the USA have been conducted by Gruber and Kubik (1997) and Autor and Duggan (2003). Both studies find that stricter screening leads to significant increases in the labor supply of older males. The study by De Jong et al. (2011) uses a field experiment to investigate the effects of intensified screening of DI benefit applications on working decisions in the Netherlands. They find that intensified screening leads to a significant decrease in both 13-week sickness absence reports and DI applications.

Among the most recent literature, Maestas et al. (2013) estimate the causal impact of DI benefit receipt on employment exploiting variation in the DI examiners' allowance rates. Using administrative data of all SSDI applicants, they find that for the applicants on the margin of program entry, the employment rate is approximately 28% points lower due to DI benefit receipt. French and Song (2014) study the impact of DI benefit receipt in the USA on labor supply taking into account the dynamics of the application process where denied individuals potentially reapply for benefits in another year. They exploit the random allocation of cases to judges and a substantial variation in judge-specific acceptance rates to show that there is a reduction in the labor supply of approximately 26% points for beneficiaries 3 years after the determination process. Moore (2015) explores a change in DI qualifying conditions in the USA related to drug and alcohol addictions in 1996 to estimate the impact of lost DI eligibility on labor force participation. The results suggest that approximately 22% of the individuals who experienced a termination of DI started working at levels higher than would actually allow them to apply for DI benefits. Their results highlight the importance of medical reassessments to justify public assistance. Frutos and Vall Castello (2015) estimate a recursive bivariate probit model for the probabilities of working and receiving DI benefits using the percentage of individuals receiving DI benefits by age group and gender as exclusion restriction. They find an average 5% points decrease in the probability of working after receiving DI benefits.

Most closely related to our paper are the studies by Chen and Van der Klaauw (2008) and Marie and Vall Castello (2012). Chen and Van der Klaauw (2008) exploit a special feature of the US system: Both the SSDI and SSI programs base disability determination decisions for some individuals not solely on medical grounds but also on vocational factors, such as age, education and work experience. They use the fact that the award rate (the probability of receiving benefits) is a discontinuous function of the age of an applicant at known age thresholds. The rationale for a fuzzy RD approach in this context is that individuals just below the cutoff age can be expected to be fairly similar to individuals just above the cutoff age in terms of their observed and unobserved characteristics but differ in terms of their DI status. Therefore, comparing individuals just below and above the cutoff age with respect to their LFP reveals the causal effect of DI benefit receipt. Overall, Chen and van der Klaauw find some evidence for distorting effects: The LFP rate of beneficiaries would have been at most 20% points higher had they not received benefits. Marie and Vall Castello (2012) exploit a similar design in the Spanish DI system. The age threshold in Spain is such that individuals above age 55 can benefit from increased DI generosity, while those below 55 are excluded. RD results reveal that increased generosity significantly reduces the labor market participation of the affected claimants.

The studies of Chen and Van der Klaauw (2008) and Marie and Vall Castello (2012) differ from our work in at least three important aspects. First and foremost, they look at the binary decision of working versus not working or the hours of work for those who do work, while we distinguish among working full-time, working part-time or being out of the labor force. This multinomial outcome allows us to investigate the impact of DI benefits on potential shifts at the extensive and intensive margins without encountering related selection issues. Second, the Swiss DI system differs from the Spanish and US systems by placing a much stronger emphasis on rehabilitation measures, and hence, we expect stronger effects at the intensive than the extensive margin. Third, we investigate effect heterogeneity to gain a better understanding of the different responses to DI benefits for different types of beneficiaries. The latter may help policy makers develop more targeted interventions in the future.

3 Institutional background

3.1 Disability insurance in Switzerland

The Swiss DI is a nationwide, compulsory social insurance that provides rehabilitation measures and cash benefits for Swiss citizens who are disabled. Similar to the SSDI program in the USA, it is mainly financed by social insurance contributions by the working population and by public funding. In 2012, the Swiss DI accounted for approximately 6.5% of the total social security expenditures in Switzerland, and it is the fourth-largest branch in the Swiss social security system, after sickness insurance (16.5%), old-age insurance (27.2%) and occupational pension plans (33.3%). Swiss citizens are eligible for benefits if there is a causal connection between the impairment to health and the inability to work. In fact, any physical or psychological health impairment, regardless of whether it is congenital, illness-related or accident-related, entitles claimants to DI benefits (Federal Social Insurance Office 2014). Furthermore, residents are only entitled to DI payments if the inability to work has lasted for at least 1 year and is likely to persist, and rehabilitation options have been exhausted.

DI is a prominent topic in the political discourse. Among the recent major reforms, the 4th revision of the Swiss DI Act in 2004 introduced regional medical screening institutions, abolished additional pensions for spouses and introduced a three-quarter pension. With the 5th revision in 2008, the leading principle of the Swiss DI was changed to "rehabilitation before pension," which broadly extended the range of options to offer disabled individuals proper incentives and support to stay in the labor market instead of depending entirely on DI benefits. Rehabilitation measures include medical measures to treat congenital disabilities, supply of appliances (e.g., wheelchairs, hearing aid devices, implants), occupational measures (e.g., career advice, retraining, vocational training, job placement, capital grants) and daily cash benefits as ancillary benefits (Federal Social Insurance Office 2013a, b). However, the 5th revision was primarily focused on measures to reduce the inflow of DI beneficiaries. Measures to reduce the stock of existing beneficiaries were only introduced with the most recent DI revision in 2012 (revision 6a), including reassessments of existing cases.

Although the Swiss DI system emphasizes reintegration, examining the DI statistics presents a different picture: Of the total expenditures of approximately 9.3 billion CHF in 2013, only 23% was spent on rehabilitation measures. The lion's share of almost 60% was used for DI pensions and helplessness allowances. An important difference between the Swiss system and the SSDI program in the USA lies in the method to calculate the amount of DI benefits. In Switzerland, the degree of disability determines the type of pension a claimant receives. The degree of disability is defined as the percentage of the loss of earnings due to disability to the potential earnings of a claimant in the absence of the impairment (henceforth earnings potential). Table 1 gives an overview of the types of pensions and minimum and maximum amounts associated with different degrees of disability. Claimants with a degree of disability of less than 40% are not entitled to pensions, whereas claimants with a degree of disability of higher than 70% are entitled to a full pension. Overall, the Swiss DI system is very generous in terms of benefits, ranking as the most generous in the world alongside the Scandinavian countries (OECD 2009).

The Swiss DI system provides rather strong incentives to take up part-time employment as the beneficiaries are allowed to increase their income above the level of the DI benefits. Specifically, recipients of a full pension are allowed to earn an additional 30% of their earnings potential and the benefits are only reduced if the earnings exceed that amount. Likewise, beneficiaries of the three-quarter (half) pension can increase their income by an extra 40% (50%) of their earnings potential. Finally, recipients of a quarter pension are allowed to earn a maximum of 60% of their earnings potential before facing a benefit cut. Based on the incentive structure in the Swiss DI system, one would naturally expect new claimants to switch, if medically possible, toward part-time employment as a reaction to the benefit receipt. Besides this first intuition on the direction of the benefit effects, the question how the benefit receipt exactly alters the labor supply decision of existing beneficiaries needs to be thoroughly analyzed empirically as the labor market participation decision is not solely determined by the benefit status.

Table 1 Degree of disability and pensions. *Source* Federal Social Insurance Office (2014)

Degree of disability	Type of pension	Minimum	Maximum
Less than 40%	No pension	–	–
40–49%	Quarter pension	277 CHF	553 CHF
50–59%	Half pension	553 CHF	1105 CHF
60–69%	Three-quarter pension	829 CHF	1658 CHF
More than 70%	Full pension	1105 CHF	2210 CHF

The minimum and maximum pensions are monthly benefits adjusted for inflation in Swiss Francs of 2007

In 2013, a total of 402,000 Swiss residents ($\approx$ 6.1% of the Swiss population) received DI services, including rehabilitation measures, DI benefits or helplessness allowances. Of these, 230,000 received cash benefits ($\approx$ 2.9% of the Swiss population) and 192,000 participated in reintegration measures ($\approx$ 2.4% of the Swiss population). Approximately 75% of the pensions were full pensions. In our data, the share of beneficiaries receiving a full pension is very close to that at approximately 72%. Furthermore, as in most Western societies, the probability of receiving disability pensions drastically increases for the group of prime-aged men and women. In numbers, approximately 60% of all recipients are aged between 40 and 64, and approximately 16% (13%) of men (women) aged 60–65 are recipients of DI benefits. From an economic point of view, these numbers raise the question whether DI is used as means of early retirement or whether the numbers simply reflect the fact that the occurrence of illness increases as part of the aging process. In fact, the first pillar of the Swiss old-age pension system, the mandatory state pension fund, allows employees to retire at most 2 years before reaching the official retirement age at the cost of lifelong reductions in the pensions. On the other hand, early withdrawals are possible from the second pillar, the mandatory occupational pension fund, starting from the age of 58, thereby effectively determining the earliest timing of retirement. Given this institutional setting, it seems plausible to presume that especially individuals with health impairments aged just below the early retirement lower bound are absorbed by the DI system until being eligible for benefits from the old-age pension system.

Approximately 15% of the DI recipients are aged 20–39. This age group primarily uses job reintegration measures offered by local DI offices. The insured below the age of 20 account for approximately 25% of the recipients. In this age segment, congenital disorders are the main cause of DI benefit receipt (Federal Social Insurance Office 2014).

3.2 Eligibility determination process

A person seeking benefits applies at the cantonal DI office. In a first step, the applicant submits medical documentation of his or her condition, as well as previous earnings records. Caseworkers at the local DI office in collaboration with an interdisciplinary team of medical doctors, specialists and vocational consultants then decide upon eligibility for DI benefits. Whether an applicant is eligible is based on a predefined set of medical and vocational factors, such as education and age, to assess the person's ability to work. Before the 4th revision of the Swiss DI system, the health assessment of the applicant was entirely based on medical certificates issued by the applicant's chosen doctor. To standardize and improve the quality of screening, the 2004 reform introduced several supra-regional medical audit institutions authorized to conduct appraisals of benefit claims and carry out medical examinations. In addition to the medical assessment, a team of vocational consultants evaluates the applicant's personal and vocational situation. The team must check for possible reintegration and rehabilitation measures reflecting the guiding principle "rehabilitation before pension." After all relevant information is gathered, the caseworkers must decide on each case within 12 months. If the decision is not accepted by the applicant, an appeal can be submitted to the cantonal insurance court within 30 days. Further levels of appeal are conducted in Federal Supreme Courts (Federal Social Insurance Office 2013a, b).

4 Data

4.1 Data source

Our data stem from the 2012 wave of the Swiss Household Panel (SHP). The SHP is especially suitable to analyze the effects of DI benefits because it offers two desirable features. First, it provides data on a stock of actual DI beneficiaries. The share of beneficiaries is approximately 3.5% on average. As in most OECD countries, this share varies by age. In line with the overall Swiss DI population, our sample consists of roughly 14–15% of the recipients aged 20–39. However, almost 80% of the beneficiaries in our sample are aged 40–64, which is higher than the official statistic, and only approximately 7.3% of the beneficiaries in our sample are aged 60–65, which is below the official statistic (Federal Social Insurance Office 2013a, b).[4] Second, unlike administrative datasets, the SHP contains rich information on numerous background characteristics, including demographic, socioeconomic and various health-related indicators. The demographic and socioeconomic status variables include the age of the respondent at the time of the interview, gender, the region of residence, the number of children living in the household between the ages of 0 and 17, an indicator for non-Swiss citizen, marital status, logarithmic gross household income in Swiss Francs, years of schooling, the person's height and weight, and an indicator for life satisfaction. The health indicators that we focus on in our analysis are the number of doctor consultations and the number of days of illness in the past 12 months, a physical activity indicator, an indicator for health impediments in everyday activities, an indicator for medication needed for everyday functioning, a dummy for self-assessed well-being and indicator variables for depression, anxiety, back problems, weakness and weariness, sleeping problems and headaches.[5] The data

[4] We discuss this data limitation in more detail in Sect. 7. While our empirical strategy identifies an internally valid local average treatment effect, the fact that the number of DI recipients in the SHP is not representative for the entire Swiss population clearly poses limitations on the external validity of our results, in particular the labor market responses of younger individuals and those directly below the retirement age.

[5] See "Appendix A.1" for a detailed description of the construction of all variables.

are restricted to the working-age population (ages 18–65) as individuals above the age of 65 are automatically transferred into the Swiss old-age pension system. Moreover, we exclude observations without accurate information on the relevant variables, such as filter errors or no answers.[6] The final estimation sample contains 3531 individuals.

4.2 DI beneficiaries versus non-beneficiaries

The endogeneity of benefit status can be characterized by comparing the subpopulations of DI beneficiaries and non-beneficiaries. Table 2 reports the mean values and differences in means for a series of relevant demographic, health and socioeconomic status indicators for both groups. The comparison of demographic characteristics reveals that the beneficiaries are on average older, have higher body weights, are lower in height and live in households with fewer children than the non-beneficiaries. Moreover, the non-recipients are more educated than the recipients, having attended school for more than one additional year on average. Additionally, non-beneficiaries typically live in households with significantly higher incomes and are more satisfied with their lives than the beneficiaries. Furthermore, the comparison of health indicators shows substantial differences between the groups, as expected. The beneficiaries visit the doctor approximately three times more often than the non-beneficiaries, and the number of days of illness for the beneficiaries is more than seven times that for the group of non-beneficiaries. Approximately 75% of the beneficiaries claim to have health impediments that severely affect their daily lives, and approximately 69% of them depend on medication to complete daily tasks. As for the medical conditions of depression, anxiety, back problems, weakness and weariness, sleeping problems and headaches, we find a significantly higher prevalence of these afflictions in the group of beneficiaries.

The bottom line from the mean comparisons is that the population of beneficiaries differs substantially from the population of non-beneficiaries, in many observable but likely also unobservable characteristics covering different health and non-health-related dimensions.[7] Any attempt to compare the groups with respect to their working decisions is therefore doomed to fail because we are not comparing similar groups that only differ in their benefit status. Revealing the isolated effect of DI benefits therefore involves at a minimum controlling for the observable characteristics, and a further exploration of the underlying mechanisms that determine DI benefit status to address unobserved confounders.

[6] Non-response is less than 1% in the SHP for the key demographic and socioeconomic variables (e.g., age, gender, education). As for the remaining health, benefit receipt and satisfaction indicators relevant for our analysis, the non-response rate is about 20% stemming mostly from vulnerable groups (e.g., individuals with migration background, low-education households) which are therefore underrepresented in the survey [see Rothenbühler and Voorpostel (2016) for details on attrition in the SHP]. Note that these more vulnerable groups can be expected to show a less accentuated labor supply reaction than the generally more flexible and literate respondents in our sample, thus leaving us with an upper bound on the DI benefit effects reported in this paper.

[7] Unobserved confounders could be, for example, genetic endowments, risk preferences, time discounting, innate abilities, work motivation and general attitudes toward health and work.

Table 2 Mean comparison of beneficiaries and non-beneficiaries

	Beneficiaries	Non-beneficiaries	Difference	*p* value
Age	50.13	44.26	5.87	0.00
Female	0.52	0.59	− 0.06	0.14
Weight (kg)	75.97	71.61	4.36	0.00
Height (cm)	169.67	170.87	− 1.20	0.12
Number of kids	0.46	0.63	− 0.16	0.04
Foreigner	0.10	0.10	0.00	0.89
Education	12.15	13.52	− 1.37	0.00
Married	0.48	0.57	− 0.10	0.02
Life satisfaction	0.66	0.89	− 0.23	0.00
Household income (CHF)	106,814	149,944	− 43,130	0.00
Doctor visits	16.05	4.71	11.34	0.00
Ill days	87.62	10.85	76.76	0.00
Physically active	0.55	0.80	− 0.26	0.00
Health impediments	0.75	0.20	0.55	0.00
Medication needed	0.69	0.20	0.49	0.00
Good health	0.45	0.84	− 0.39	0.00
Depression	0.50	0.23	0.27	0.00
Back problems	0.30	0.11	0.19	0.00
Weariness	0.32	0.11	0.22	0.00
Sleeplessness	0.20	0.09	0.12	0.00
Headaches	0.11	0.09	0.03	0.27
Number of observations	143	3388		

p values for the null hypothesis of no difference in means between beneficiaries and non-beneficiaries

5 Identification strategy

The objective of this study is to identify the effect of DI benefits on the labor market decisions of existing DI beneficiaries in Switzerland. We apply a fuzzy regression discontinuity (RD) design to overcome the endogeneity of benefit status by exploiting a discrete jump in the benefit award rate at age 56. Specifically, we estimate the effects of DI benefits on the individual decision to work part-time, work full-time or be out of the labor force using discrete endogenous switching (ES) models (Miranda and Rabe-Hesketh 2006; Roodman 2011).

5.1 Discontinuity in the benefit award rate

The benefit award rate as the probability of receiving DI benefits in relation to a person's age is depicted in Fig. 1. The dots represent the mean share of beneficiaries over bins of half-years of age. The scatter plot is overlaid with a quadratic fit and a corresponding 99% confidence interval. The dashed red vertical line indicates the age of 56. The graph shows that for the individuals below the age of 56, the benefit

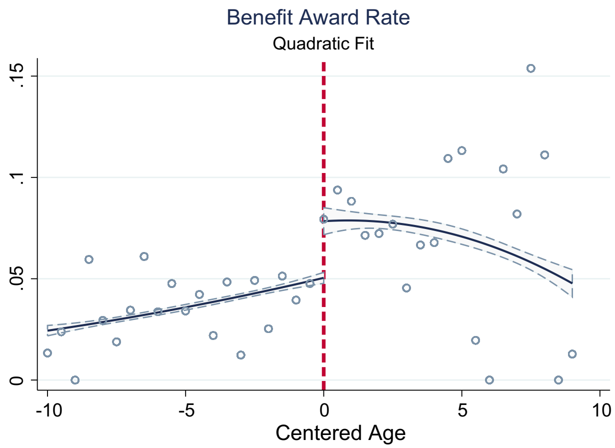

Fig. 1 Discontinuity in the benefit award rate. *Note* The figure shows the benefit award rate in half-year bins of centered age. Age is centered at 56 years. The raw means are overlaid with a quadratic fit and corresponding 99% confidence interval to the left and right of the age cutoff at 56 years

award rate is generally stable at a level of approximately 4–5%. Above the age of 56, the probability of receiving DI benefits is well above the 5% mark indicating a discontinuous jump in the benefit award rate at age 56. This finding means that individuals just below that age cutoff have a significantly lower probability of receiving DI benefits than people just above the discontinuity. The discontinuity is also confirmed when running classical first-stage regressions using the benefit status as the dependent variable. Table 3 shows the first-stage effects for four different model specifications. In all cases, the discontinuity is estimated to be approximately 4–5% points, which coincides with the graphical evidence in Fig. 1. The Cragg–Donald Wald F-statistics are above the usual critical values rejecting the null of a weak instrument and therefore indicating instrument relevance (Stock and Yogo 2005).

Why do we observe a discontinuity in the benefit award rate? As in most OECD countries, the eligibility determination process in Switzerland is explicitly based on both medical and vocational factors, including the social situation and age of the applicant (Federal Social Insurance Office 2013a, b). The fact that local DI offices take into account such background characteristics when deciding upon DI benefits leads to differences in the likelihood of receiving DI benefits for different subgroups and characteristics of applicants. For example, individuals losing their job at an older age tend to have more problems reentering the job market than younger individuals. This is particularly true for unskilled workers where job reintegration measures are usually not a valid option. In such cases, DI is often used as a substitute for early retirement. This mechanism and age 56 as a critical threshold have been confirmed by several heads of DI departments in personal conversations, which in turn explains the sharp increase in the share of beneficiaries at this particular age threshold. In related evidence, Chen and Van der Klaauw (2008) show that the age of the DI applicant is also a key factor in the benefit determination screening process in the USA. Staubli (2011) uses changes in the eligibility criteria above certain age thresholds to identify

Table 3 First-stage regressions of DI benefit receipt on age threshold

Dependent variable: DI benefits	(1)	(2)	(3)	(4)
I (age ≥ 56)	0.047***	0.041**	0.049**	0.046**
	(0.02)	(0.019)	(0.021)	(0.019)
Control variables	No	Yes	No	Yes
Quadratic age function	No	No	Yes	Yes
Number of observations	3531	3531	3531	3531
R^2	0.05	0.166	0.05	0.22
Cragg–Donald Wald F-stats	7.04	6.87	6.32	6.98

First-stage linear regression results using the indicator for DI benefits as dependent variable, a standard RD specification and a data window of age between 35 and 65. The instrumental variable is the indicator for age 56 or above, and all specifications include age centered around the threshold and an interaction of centered age and the instrument. Controls include the demographic, socioeconomic and health-related variables as described in Sect. 4. Standard errors are clustered at the household level
Significance levels: $^{***}p < 0.01$; $^{**}\ p < 0.05$; $^{*}p < 0.1$

the impacts of DI on employment, and Marie and Vall Castello (2012) explore age thresholds similar to our strategy to measure the impacts of DI generosity on LFP in Spain.

5.2 RD validity checks

Identification in the RD framework is based on the individuals' inability to precisely control the assignment variable (in our case age) near the threshold. Given the timing and sampling scheme of the SHP and the modalities of the DI application process, which accounts for a variety of factors other than age, such as employment history and family background, we deem it unlikely that the RD identification assumption is violated. Under this presumption, individuals just below the age threshold of 56 will form a natural control group for those above it, and observable as well as unobservable background characteristics should be balanced around the cutoff such that the treatment (DI benefit receipt) is "as good as randomized" (Lee and Lemieux 2010).

Since individuals cannot manipulate their age, it is reasonable to assume that they have no precise control over whether they are observed just to the left or to the right of the age cutoff. However, there is also a potential threat to this identifying assumption. DI applicants can influence the timing of their applications, e.g., with the intention of maximizing their chances of receiving DI benefits. To assess the validity of the RD design, we therefore check for discontinuities in the number of DI applications and for discontinuities in baseline covariates at the age cutoff. Both discontinuities in the number of DI applications and predetermined variables would cast doubt on the validity of the RD design and thus our identification strategy.

The rationale behind the first test (discontinuity in the number of DI applications) is that if the age threshold at age 56 was known by DI applicants, then weak cases

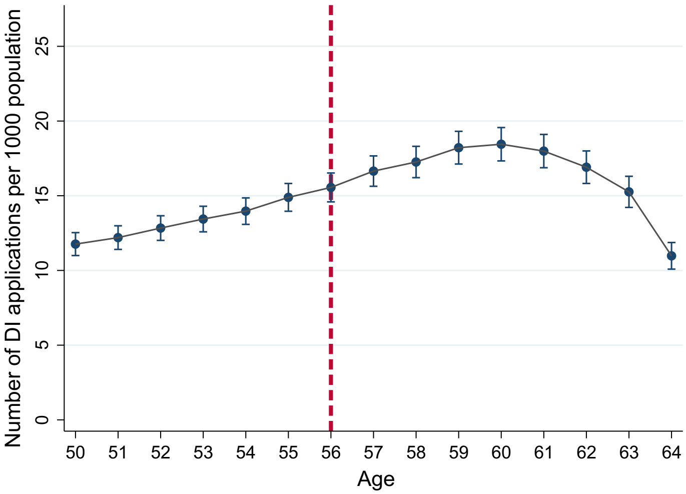

Fig. 2 Number of DI applications per 1000 population by age. *Note* The figure shows the overall number of disability insurance applications per 1000 population for individuals aged 50–64

aged 55 (or lower) would have a strong incentive to postpone their application by another year and therefore increase the likelihood of being awarded DI benefits.[8] In order to test this claim, we use administrative data provided by the Federal Social Insurance Office (2016) on the total number of DI applications per 1000 population for individuals aged 50–64 to check for potential manipulation at age 56. Figure 2 shows a positive trend in the number of DI applications reaching a maximum of roughly 18 applications per 1000 population at the age of 60. After the age of 60, the number of applications drops as individuals are gradually transferred into the Swiss old-age pension scheme. DI beneficiaries are automatically eligible for old-age pensions (and loose eligibility for DI benefits) once they reach the official retirement age of 65 for men and 64 for women. As touched upon in Sect. 3.1, there are two possibilities to retire early in Switzerland: First, employees are legally allowed to retire before the official retirement age upon reaching the age of 58 at the earliest, in exchange for a lifelong reduction in their pension.[9] Second, starting from the age of 58, partial retirement is possible if employees continue to earn at least 50% of their former salary with the benefit of not suffering any pension reductions. Approximately 58% of the Swiss working population retires before reaching the official retirement age, thereby loosing the eligibility to collect DI benefits, which explains the decrease in DI applications around the age of 60.

Of greater importance for our identification strategy, however, is the fact that there are no signs of a discontinuity in the number of DI applications at the age of 56, and therefore, we do not find evidence from the official statistics on DI applications for

[8] Note that the standard procedure in the RD literature of investigating the density of the forcing variable around the threshold does not work in our case because of the sampling design, which is a representative sample of the whole Swiss residential population, including all age groups.

[9] Pension reductions currently amount to approximately 5–8% for every additional year of early retirement.

a manipulation at the age cutoff in terms of postponing DI applications around this critical age.

As a second validity check, we compare observable characteristics over the age range to determine whether they are locally balanced around the age cutoff. In fact, local random assignment implies that both observable and unobservable factors should not systematically differ between people below and above the cutoff. In Fig. 3, we show the share of parents with medium-to-low levels of education,[10] own years of schooling, the share of women, the number of kids and a series of health indicators in a 10-year window around the age cutoff. The graph indicates no statistically significant jump for any of these covariates at the age cutoff. In an unreported analysis, we also look at the pattern in some of the other variables listed in Table 2 at the age threshold. While we do not find significant mean differences from just below to above the cutoff for any of these variables, we do not want to overemphasize this result because some of the variables, such as reported health status or physical activity, may be causally linked to DI benefit status.

In addition, we test whether alternative age cutoffs lead to a change in the DI benefit award rate. This finding would indicate that our results are driven by age effects instead of effects linked to DI benefit receipt. Table 4 shows alternative first-stage regressions for the age thresholds 52, 54, 59 and 61 using the same specification as in Table 3, column 2. None of the alternative age cutoffs turns out to be significant, supporting the claim that the obtained discontinuity at age 56 is systematic to DI practice instead of general aging. Thus, overall, the different checks do not provide evidence against the validity of the RD identifying assumption.

5.3 The discrete ES model

A discrete ES model is used to quantify the incentive effects of DI benefits on the working decisions of existing beneficiaries. The theoretical model is derived using the framework of additive random utility models (ARUM) providing a natural connection to economic choice theory (Marschak 1960). It is assumed that each decision maker i faces three working choices: $y_i \in \{$working part-time $(j = 1)$, working full-time $(j = 2)$ or being out of the labor force $(j = 3)\}$. Furthermore, agents are assumed to be rational in the sense that they choose the alternative that is associated with the highest utility for them. Utility is known to the decision maker but not to the researcher. The researcher only observes some attributes of the agent and the final working decision that the individual makes (Train 2009). The utility for individual i from option j is specified as

$$U_{ij} = x_i'\beta_j + D_i\gamma_j + \varepsilon_{ij} \quad \forall \quad j = 1, 2, 3;\ n = 1, ..., N \tag{1}$$

where x_i is a vector of observable characteristics, including the health indicators, demographic and socioeconomic factors described in Sect. 4 and a flexible function in age to account for the RD setup; $D_i \in \{0, 1\}$ is the binary treatment indicator

[10] Parental education is considered medium to low if the educational attainment of the parents was at most compulsory schooling or having finished an apprenticeship (upper secondary education).

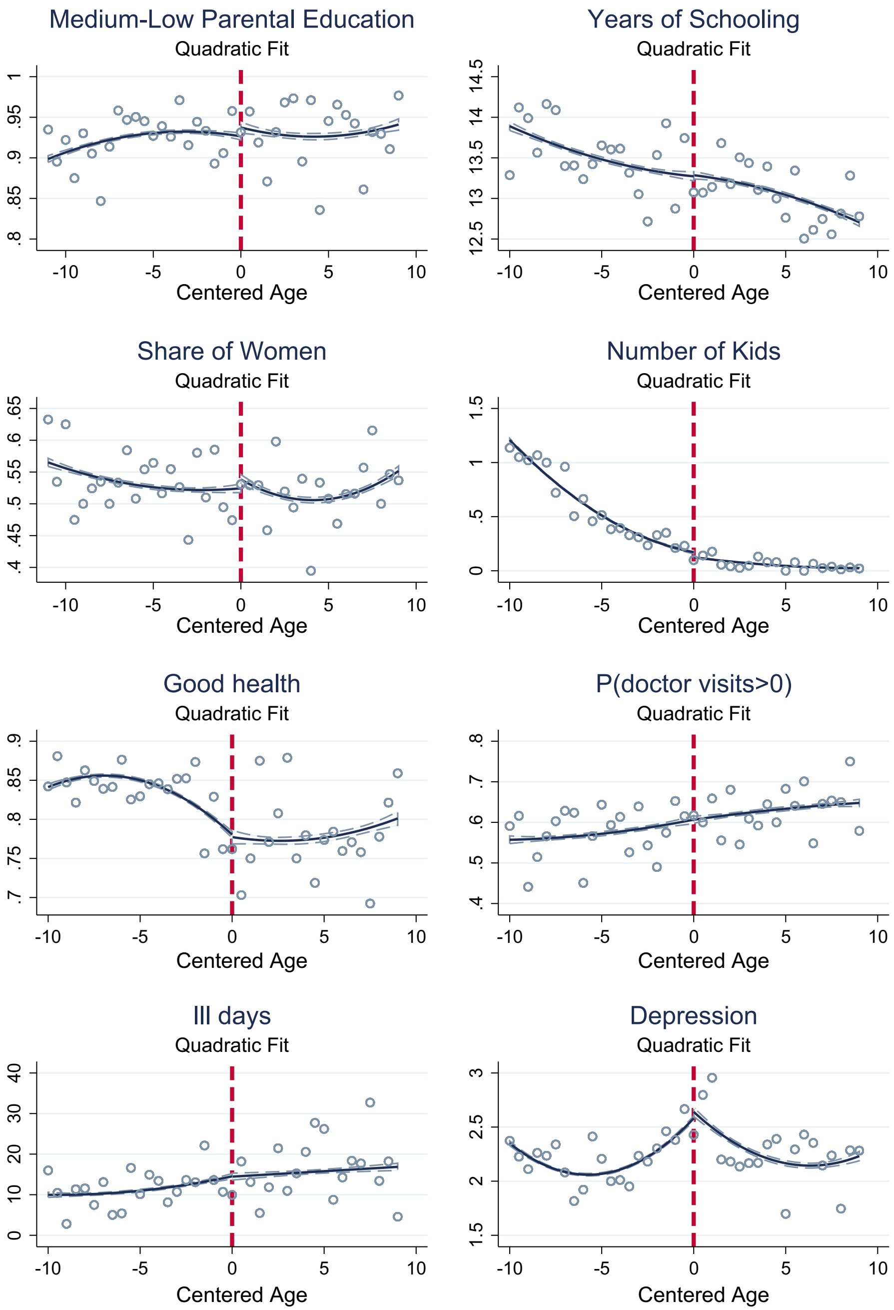

Fig. 3 Inspection of baseline covariates. *Note* The figure shows a selection of baseline variables in a window of ± 10 years around the age cutoff overlaid with a quadratic fit and a corresponding 99% confidence interval

Table 4 Test of alternative age cutoffs

Dependent variable: DI benefits	(1)	(2)
$I(\text{age} \geq 52)$	− 0.02	− 0.01
	(0.02)	(0.01)
$I(\text{age} \geq 54)$	0.01	− 0.00
	(0.02)	(0.01)
$I(\text{age} \geq 59)$	0.02	0.02
	(0.03)	(0.02)
$I(\text{age} \geq 61)$	0.02	− 0.00
	(0.03)	(0.03)
Control variables	No	Yes
Number of observations	3531	3531

First-stage linear regression results using the indicator for DI benefits as dependent variable, a standard RD specification and a data window of age between 35 and 65. Instruments are as defined in the table. All specifications include age centered around the threshold and an interaction of centered age and the instrument. Controls include the demographic, socioeconomic and health-related variables as described in Sect. 4. Standard errors are clustered at the household level
Significance levels: ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

for DI benefit receipt; $\theta_j \equiv \{\beta_j, \gamma_j\}$ is the vector of model parameters with γ_j of main interest; and ε_{ij} is an unobserved error term that includes choice-relevant factors not contained in x_i. A binary threshold crossing model is introduced to address the endogeneity of benefit status:

$$D_i^* = w_i'\delta + v_i \quad (i = 1, ..., N)$$
$$D_i = \begin{cases} 1 & \text{if } D_i^* > 0 \\ 0 & \text{if } D_i^* \leq 0 \end{cases} \tag{2}$$

where D_i^* is a continuous random variable reflecting the latent propensity to receive DI benefits as a function of observable variables w_i, which are the same as in x_i but also include the age cutoff as instrumental variable for the benefit status in Eq. (1); δ is a vector of parameters and v_i an error term in the DI benefit status equation.

The vector ψ_i is composed of all the error terms from Eqs. (1) and (2) and is assumed to follow a multivariate normal distribution with a mean vector of zero and covariance matrix Ω:

$$\psi_i = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ v_i \end{pmatrix} \sim \mathcal{MVN}(\mathbf{0}, \mathbf{\Omega}) \qquad \text{where } \Omega = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{1v} \\ \cdot & \sigma_2^2 & \sigma_{23} & \sigma_{2v} \\ \cdot & \cdot & \sigma_3^2 & \sigma_{3v} \\ \cdot & \cdot & \cdot & \sigma_v^2 \end{pmatrix} \tag{3}$$

To ensure parameter identification, one must acknowledge that the level and scale of the latent utilities are irrelevant in the model. The absolute level of utility is irrelevant because adding any constant c to the utility of each working option does not change the ordering of utilities and therefore has no effect on the observed labor supply choice of the beneficiary. The level of utility is normalized by choosing a base category (we will use part-time as the base throughout the analysis) and setting the corresponding parameter vector to zero, i.e., $\theta_1 = 0$. Similarly, the scale is irrelevant because each latent utility can be multiplied by a positive constant c without changing which alternative has the highest utility (Train 2009). The scale is normalized by imposing constraints on Ω, as discussed by Bunch (1991) and Roodman (2011).[11]

5.4 Simulation of choice probabilities

Using the structural form of the model, probabilistic statements about individual work decisions can be made. For example, the probability of working full-time given DI benefits is given by

$$\begin{aligned}
P(y_i = 2|D_i = 1, x_i) &= \frac{P(U_{i2} > U_{i1}, U_{i2} > U_{i3}, D_i^* > 0)}{P(D_i^* > 0)} \\
&= \int_{-\infty}^{x_i'\beta_2 + D_i\gamma_2} \int_{-\infty}^{x_i'(\beta_2-\beta_3)+D_i(\gamma_2-\gamma_3)} \phi_2(\tilde{\varepsilon}_{12}, \tilde{\varepsilon}_{32}) d\tilde{\varepsilon}_{12} d\tilde{\varepsilon}_{32} \\
&\quad - \frac{\int_{-\infty}^{x_i'\beta_2} \int_{-\infty}^{x_i'(\beta_2-\beta_3)} \int_{-\infty}^{-w_i\delta} \phi_3(\tilde{\varepsilon}_{12}, \tilde{\varepsilon}_{32}, v_i) d\tilde{\varepsilon}_{12} d\tilde{\varepsilon}_{32} dv_i}{\Phi(-w_i\delta)} \\
&= \Phi_2(x_i'\beta_2 + D_i\gamma_2, x_i'(\beta_2 - \beta_3) + D_i(\gamma_2 - \gamma_3); V_1) \\
&\quad - \frac{\Phi_3(x_i'\beta_2, x_i'(\beta_2 - \beta_3), -w_i\delta; V)}{\Phi(-w_i\delta)}
\end{aligned} \tag{4}$$

where $\phi_2(\cdot)$ and $\phi_3(\cdot)$ are bi- and trivariate standard normal densities, $\Phi_2(\cdot)$ and $\Phi_3(\cdot)$ are the corresponding bi- and trivariate cumulative density functions with correlation matrix V of the differenced errors $\tilde{\varepsilon}_{jk} = \varepsilon_j - \varepsilon_k$ in the utility equations for options j and k, and $V_1 \equiv \rho_{\tilde{\varepsilon}_{12},\tilde{\varepsilon}_{23}}$ is the correlation between $\tilde{\varepsilon}_{12}$ and $\tilde{\varepsilon}_{23}$.

Equation (4) demonstrates that the choice probabilities in the discrete ES model are multivariate integrals over subsets of the Euclidean space. The problem is that these choice probabilities cannot be expressed in closed form. Instead, one must use simulation methods to evaluate the integrals numerically. We use the Geweke, Hajivassiliou and Keane (GHK) algorithm (Geweke 1991; Hajivassiliou and McFadden 1998; Keane 1990, 1994), which has been shown to simulate normal probabilities well (Hajivassiliou et al. 1994). The GHK algorithm is based on the idea that expressions such as (4) can be rewritten as a sequence of conditional probabilities that can be simulated recursively. The algorithm takes a predefined number of draws from the unit interval for each observation and then generates the simulated probability at each iteration step. We use the GHK-based conditional mixed process estimator as it is programmed in the Stata routine `cmp` by Roodman (2011). For each of the estimated models, we use 1,000 random draws per observation to simulate the choice probabilities. Following

[11] We follow the standard approach in the literature by imposing variance unity on the error terms to normalize the scale of utility and thus ensure parameter identification.

Train (2009), antithetic Halton draws are used to maximize uniformity of coverage of the unit interval and to achieve greater accuracy than with pseudorandom draws. The latent model parameters are estimated by maximum simulated likelihood (MSL); the corresponding log-likelihood function is shown in "Appendix A.2." The appendix also presents the results for varying numbers of simulation draws confirming that our findings are stable for the chosen setup.

5.5 Treatment effects

To illustrate the DI benefit effects, we use the parametric model structure to derive the treatment effect on the treated (TOT) for each labor market outcome. To give an example, the TOT for the outcome of working full-time is given by

$$\begin{aligned} \mathrm{TOT}_{\mathrm{full}} &= P(y_{i1} = 2|D_i = 1, x_i) - P(y_{i0} = 2|D_i = 1, x_i) \\ &= \Phi_2(x_i'\beta_2 + D_i\gamma_2, x_i'(\beta_2 - \beta_3) \\ &\quad + D_i(\gamma_2 - \gamma_3); V_1) - \Phi_2(x_i'\beta_2, x_i'(\beta_2 - \beta_3); V_1) \end{aligned} \tag{5}$$

and similarly for the TOT of working part-time and being out of the labor force. The TOT reflects the difference between the probability of working full-time given that beneficiary i received DI benefits and the probability of working full-time given that the same beneficiary did not receive benefits. A negative TOT therefore indicates that DI benefit receipt decreases the probability of working full-time and vice versa for a positive TOT. The TOTs can be calculated for each individual, averaged over individuals, and we can look at the distribution of TOTs. The TOTs are estimated using the MSL estimates of the parameters, based on the discrete ES model.

6 Results

The results are presented in two steps. First, we discuss the impact of DI benefit receipt on the labor market decision for the average beneficiary. Specifically, we show the average TOTs, as well as the coefficient estimates from the discrete ES model, for two model specifications using different sets of control variables. Second, we further investigate the incentives inherent in DI benefits by showing the entire treatment effect distributions for each of the three labor market outcomes. We then compute the quartiles of each of these distributions to split the beneficiaries into four equally sized groups. Isolating the four groups allows us to characterize effect heterogeneity by looking at mean values of background variables within each group.

6.1 Average DI benefit effects

The estimated coefficients of the discrete ES model are displayed in Table 5. The table is organized such that the upper panel shows the coefficients on the DI benefits indicator for the outcomes of working full-time and being out of the labor force (part-time is the base category) and the lower panel shows the coefficients on the instrument

Table 5 Discrete ES estimates of the effect of DI benefits on labor supply

	(1)	(2)
Panel A: Outcome equations		
Working full-time		
DI benefits	− 2.05***	− 2.04*
	(0.57)	(1.21)
Out of labor force		
DI benefits	− 0.64	− 0.04
	(0.86)	(1.17)
Panel B: First-stage equation		
$I(age \geq 56)$	0.31**	0.43**
	(0.13)	(0.19)
Background 1: Demographics	Yes	Yes
Background 2: SES and health	No	Yes
Number of observations	3531	3531
Number of iterations	9	9
Pseudo-log-likelihood	− 2203.98	− 1794.72
Computation time (s)	966.41	1106.55

Discrete ES estimates of the coefficient on DI benefit status for outcomes working full-time and out of labor force (base category: part-time) based on 1000 antithetic GHK draws. Coefficients on background variables not shown. Background 1: age centered around the threshold and interaction of centered age and the instrument, gender, weight, height, number of kids, foreigner, dummies for region of living and type of community. Background 2: marital status, years of schooling, log of household income, life satisfaction, number of doctor visits, number of ill days, physical activity, health impediments, medication needed, indication for self-assessed health and dummies for depression, back problems, weariness, headaches and sleeping problems. Standard errors are clustered at the household level

Significance levels: ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

in the DI benefits equation. In both models, the discontinuity in the benefit award rate above the age of 56, i.e., an indicator for crossing the age threshold, is used as the instrument for DI benefit status. In the first model (column 1), we include control variables for individuals' demographic background.[12] In this model, we obtain a highly significant and negative coefficient on the DI benefit variable regarding the latent utility of working full-time but no statistically significant coefficient on DI benefits regarding the latent utility of being out of the labor force. In the second model (column 2), we add control variables for socioeconomic background[13] and health status.[14] Again, we find a statistically significant and negative coefficient on DI benefits of

[12] Age, age interacted with the instrument, gender, weight, height, number of children, foreigner status and six regional dummies according to the standard classification of the Swiss Federal Office of Statistics.

[13] Marital status, years of schooling, log household income, life satisfaction.

[14] Number of doctor visits, days of illness, physical activity, health impediments, medication needed, indication for self-assessed health and dummies for depression, back problems, weariness, headaches, sleeping problems.

Table 6 Average effects of DI benefits on labor market outcomes

	Mean	SD	Minimum	Maximum
Working full-time	− 0.35	0.28	− 0.77	− 0.00
Working part-time	0.32	0.22	0.01	0.69
Out of labor force	0.04	0.07	− 0.01	0.40

Summary statistics of the TOTs for the labor market outcomes working full-time, part-time and out of labor force. The TOTs are based on the coefficient estimates from the discrete ES model as it is shown in Table 5, column (2)

approximately the same size for working full-time. The estimated coefficient on DI benefits for being out of the labor force is close to zero and not statistically significant. The discrete ES model also produces a highly significant first-stage effect, which is of same magnitude as that shown in Table 3. While the coefficients are not directly comparable between the models due to the different model structures, the average marginal probability effect for the instrument can be translated into a 4.5% points effect for both specifications. Note that the estimates in the second specification are basically unchanged if further controls for the number of hospital days and parental education as a proxy for genetic endowment are included, reinforcing the robustness of the presented findings.

The negative and significant coefficient on DI benefit status for the latent utility of working full-time suggests, at first glance, that DI benefits provide strong work disincentives. However, to appropriately address the question of the impact of DI benefit receipt on the discrete labor market outcome of working full-time, working part-time or being out of the labor force, one must translate the coefficients into probability effects as outlined in the previous section. The TOTs provide the relevant information on changes in the labor supply decision of an existing beneficiary to the working decision of the same person in the absence of DI benefits (the counterfactual). Table 6 shows summary statistics (means, standard deviations, minimum and maximum values) of the TOTs for each of the three labor market outcomes.

Beginning with the effects of DI benefits on the decision to work full-time, we see that the probability of working full-time decreases on average by approximately 35% points, or in other words, the probability of working full-time would be 35% points higher for the average beneficiary if he or she were not receiving DI benefits. All the estimated TOTs are strictly negative (by model structure), ranging from approximately 0% points to a minimum of − 77% points. On the other hand, receiving DI benefits on average increases the probability of working part-time by approximately 32% points, i.e., the probability of working part-time would be 32% points lower for the average beneficiary if he or she were not receiving DI benefits. The TOTs are strictly positive, ranging from approximately 0% points to + 69% points. The results for the probability of being out of the labor force indicate a small increase by approximately 4% points for the average beneficiary. However, as for the other two outcomes, there is substantial effect heterogeneity, with some beneficiaries affected negatively and some by as much as + 40% points. This heterogeneity will be further analyzed in the next section.

6.2 Effect Heterogeneity

To gain further insight into the effects of DI benefits on the labor supply of existing beneficiaries, we explore the TOT distributions for each of the three labor market outcomes. The top histogram in Fig. 4 shows the distribution of TOTs for the probability of working full-time, including the corresponding quartiles dividing the sample into four equally sized groups. The TOTs are weakly negative for all beneficiaries. The probability of working full-time for beneficiaries in the lowest quartile of the TOT distribution is reduced by between 67% points and 77% points. In contrast, for beneficiaries in the upper quartile the probability of working full-time is reduced by between 0% points and 9% points. When we compare the beneficiaries in the different TOT quartiles regarding their background characteristics (Table 7), those in the lower quartile with a strong negative reaction to DI benefits are mostly men with a comparably high socioeconomic status in relatively good health. However, people in poor health, with a low socioeconomic status, and women tend to adjust their probability of working full-time only by a small amount. Of course, the predicted probability of working full-time at baseline, i.e., without DI benefits, for the latter group is also lower than for the former group (3 vs 84% on average). Thus, while the absolute effect of DI benefits on the probability of working full-time is largest for male individuals, with a relatively high socioeconomic status and in good health, the relative effect is larger for those groups that tend to be more disadvantaged in the labor market (i.e., women, people with low socioeconomic status and ill individuals).

Regarding part-time work, the middle histogram in Fig. 4 provides two key insights. First, the TOTs are all nonnegative, which indicates that DI benefit receipt increases the probability of working part-time for all the beneficiaries in the sample. Second, the TOTs range from 0% points to a maximum of 69% points, which suggests substantial effect heterogeneity between DI beneficiaries. In fact, the lower quartile of the TOTs ranges from 0% points to 12.6% points, suggesting small to modest reactions in the labor supply decision as a response to benefit receipt. In contrast, beneficiaries in the upper quartile respond dramatically to the benefits, with the likelihood of working part-time increasing by approximately 53% points up to 68% points. Comparing the beneficiaries in the highest quartile with the strongest reaction to DI benefit receipt to those in the lowest quartile with the weakest reaction, three major differences between the groups can be identified. The reaction to DI benefits is strongest for (i) men who are in (ii) relatively good health, as indicated by the lowest numbers of doctor visits and days of illness, and have (iii) relatively high incomes. Their baseline probability of working part-time without DI benefit receipt is approximately 15.5%. The fact that particularly those beneficiaries with a low income react the least to the benefits can be explained by the comparably high replacement rate of benefits for those individuals. Additionally, being in poor health might explain a general inability to adjust the individual labor supply, with many individuals in poor health already working part-time (approximately 79%) or being out of the labor force (9%).

Finally, the bottom histogram of Fig. 4 presents the distribution of TOTs for the probability of being out of the labor force. This distribution has the highest mass around small positive effects; however, the upper quartile of the distribution indicates

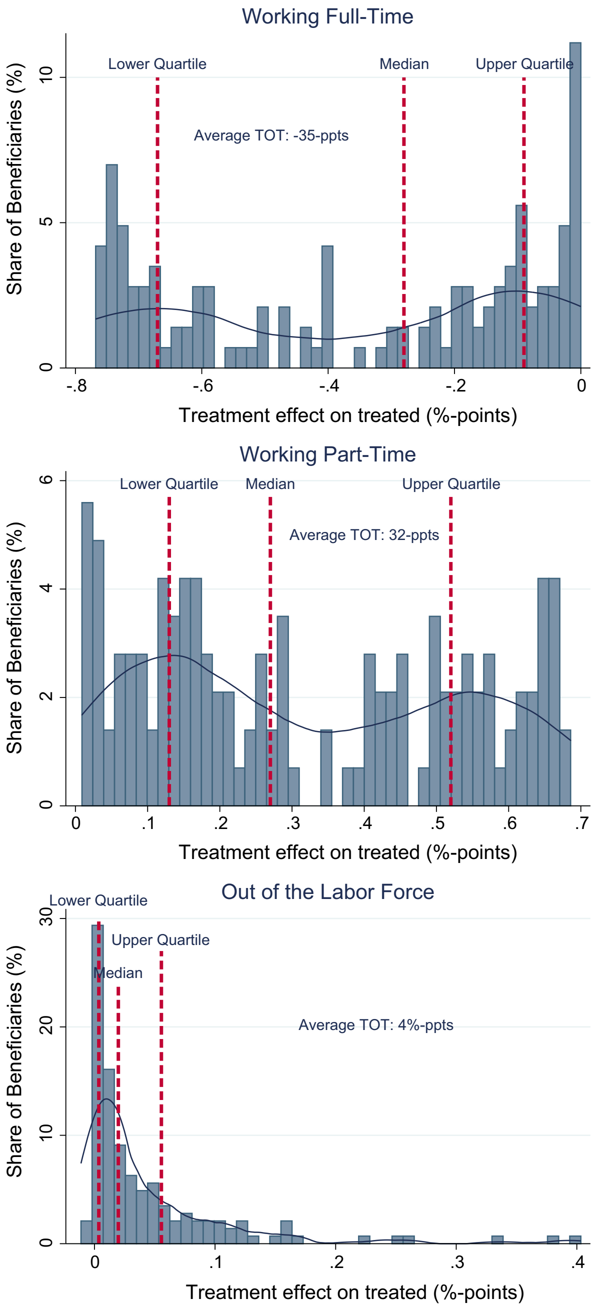

Fig. 4 Distributions of the treatment effects on the treated. *Note* The figure shows the distributions of the individual treatment effects on treated (TOT) for the labor market outcomes of working full-time, part-time and being out of the labor force, overlaid with a kernel density estimate. The dashed lines indicate the 25th, the 50th and the 75th percentiles. A positive (negative) sign on the TOT indicates an increase (decrease) in the probability of observing each of the outcomes as a response to the DI benefit receipt in the population of beneficiaries

Table 7 Mean values of background characteristics by TOT quartile

Working full-time	Lower quartile [– 0.768, – 0.674]	25–50th percentile (– 0.674, – 0.283]	50–75th percentile (– 0.283, – 0.090]	Upper quartile (– 0.090, – 0.000]
Number of doctor visits	13.5	14.4	8.8	27.9
Health status	3.2	3.5	3.6	3.0
Number of ill days	121.3	65.9	51.9	112.5
Age	53.2	47.7	45.9	53.9
Female	0.0	0.3	0.8	1.0
Years of schooling	12.8	12.5	11.5	11.8
Household income (CHF)	111,765	111,696	113,423	89,849
Working part-time	Lower quartile [0.009, 0.126]	25–50th percentile (0.126, 0.268]	50–75th percentile (0.268, 0.524]	Upper quartile (0.524, 0.686]
Number of doctor visits	26.4	10.9	11.8	15.0
Health status	3.0	3.5	3.3	3.5
Number of ill days	100.7	86.1	91.4	71.8
Age	52.6	47.3	47.8	52.9
Female	0.9	0.7	0.4	0.1
Years of schooling	11.8	11.6	12.3	12.9
Household income (CHF)	100,841	97,861	101,428	112,770
Out of labor force	Lower quartile [– 0.011, 0.0036]	25–50th percentile (0.0036, 0.019]	50–75th percentile (0.019, 0.055]	Upper quartile (0.055,0.403]
Number of doctor visits	17.4	12.4	15.6	18.9
Health status	3.8	3.5	3.1	2.9

Table 7 continued

Out of labor force	Lower quartile [− 0.011, 0.0036]	25–50th percentile (0.0036, 0.019]	50–75th percentile (0.019, 0.055]	Upper quartile (0.055,0.403]
Number of ill days	36.4	36.3	106	172.2
Age	43.1	50.9	53.3	53.0
Female	0.3	0.6	0.7	0.4
Years of schooling	12.4	12.1	12.1	12.1
Household income (CHF)	148,990	105,951	98,636	74,168

Range of TOTs in quartiles for each labor market outcome are indicated in column heads. Mean values of selected variables for existing beneficiaries are shown for each of the four quartiles of the TOT distribution for each outcome (working full-time, working part-time and out of the labor force) effects of DI benefits from 5.5 to 40.3% points. Consistent with the above results, Table 7 indicates that it is mainly the low-income individuals and beneficiaries in poor health who drop out of the labor market. In contrast, the better off beneficiaries show essentially no response to the benefit receipt regarding the decision of dropping out of the labor market.

7 Discussion and conclusion

Over the past few decades, the number of beneficiaries and the costs of DI programs have increased sizably in most OECD countries, which poses a major challenge for the future financing of social security systems. At the same time, only a small fraction of those entitled to benefits ever leave the DI system, although some might have valuable work capacities. Policy makers who seek to avoid losing these resources and effectively improve the incentive structure of the DI system therefore must ask the question of how working decisions of existing beneficiaries are affected by the provision of DI benefits. We investigate this question by analyzing potential lock-in effects created by the financial incentives embodied in the Swiss DI system.

Our empirical strategy is based on a fuzzy RD design that exploits a discontinuity in the DI benefit award rate to assess individual benefit status. Allowing for a multinomial outcome of working full-time, working part-time or being out of the labor force, we estimate a discrete ES model (Miranda and Rabe-Hesketh 2006; Roodman 2011) to identify the effects of DI benefits on the labor market decisions of existing beneficiaries. Unlike the existing literature (e.g., Gruber 2000; Chen and Van der Klaauw 2008; Marie and Vall Castello 2012; French and Song 2014; Moore 2015), we do not confine ourselves to the binary choice of working versus not working. The main reason is that we expect stronger effects at the intensive than at the extensive margin due to the particularities of the Swiss DI system, with beneficiaries more likely shifting their labor supply from full-time to part-time than dropping out of the labor force. Moreover, by looking at the multinomial labor market outcome, we address a more general point on possible selection issues when investigating the impact of DI benefits on the intensive margin. These issues arise because the impact of DI benefits on the hours of work may be confounded by the impact on the extensive margin (see Angrist and Pischke 2009). On the downside, our approach comes at the cost of a parametric model structure for the multinomial labor market outcome and endogenous DI benefit equations. Although this might make the analysis prone to misspecification, the model can be well justified by random utility maximization, and choice modeling in general is most often based on parametric simulation approaches (Train 2009).

Conditional on health status and demographic and socioeconomic background, the results of the discrete ES model suggest that DI benefit receipt strongly influences the labor market decisions of existing beneficiaries. DI benefits affect the average beneficiary mainly at the intensive margin, with a shift in the work intensity from full-time to part-time, whereas the probability of being out of the labor force is hardly influenced. The latter result contrasts with some of the earlier literature, which has mainly focused on the extensive margin (e.g., Bound 1989; Gruber 2000; Marie and Vall Castello 2012; Frutos and Vall Castello 2015). However, our results must be interpreted in the light of the Swiss DI system, which on the one hand emphasizes work integration before pension and on the other hand allows for partial DI benefits. The few existing studies that look at the impact of features of the Swiss DI system on the labor supply (e.g., Kauer 2014; Bütler et al. 2015; Eugster and Deuchert 2015) are consistent with our result of small or no effects at the extensive margin and some effects at the intensive margin, although the earlier evidence for the effects at the intensive margin must be interpreted with care due to the possible selection issues in the conditional-on-positives samples.

Our results indicate substantial heterogeneity in the effects of DI benefits on labor market outcomes, i.e., different groups of the population are affected very differently by DI benefit receipt. We find that men are more likely to shift their labor supply from full-time to part-time, whereas women tend to drop out of the labor force. Moreover, beneficiaries living in middle- to high-income households and who are in relatively good health tend to shift their labor supply from full-time to part-time. In contrast, low-income and relatively ill individuals tend to drop out of the labor force. A positive interpretation of these findings would be that the DI system works in the sense that people who receive benefits remain in the work force as long as they are capable. A more critical interpretation is that the financial incentives provided by the DI system trigger beneficiaries to lower their work intensity and therefore keep them dependent on income transfers (see also Kostol and Mogstad 2014, for related arguments in the context of the DI system in Norway). This in turn adds a possible explanation for the low DI outflow in Switzerland. Our results help clarify the underlying mechanisms explaining DI outflow by characterizing the heterogeneity in reactions to DI benefit receipt, which should support policy makers to better target policy interventions to certain groups of beneficiaries.

Finally, this paper has several limitations, opening up avenues for future research: First, the analysis is based on a fairly low number of beneficiaries due to data limitations in the SHP that cannot be overcome by the researcher. Moreover, since the age structure of DI beneficiaries in our sample does not perfectly match with the one reported in the official DI statistics for the year 2012, the reported labor supply effects in this paper have to be interpreted with caution and cannot simply be extrapolated to the population of DI beneficiaries in Switzerland. It would therefore be important to see future research revisiting the question on the employment effects of DI benefits based on a larger and more representative sample of recipients. Second, although our analysis provides prime evidence for substantial heterogeneity in the treatment effects across different groups of beneficiaries, it would be interesting to see future research further exploring effect heterogeneity, for example by the degree and/or type of disability.

Acknowledgements We are grateful to Magne Mogstad, Judit Vall Castello, Teresa Bago d'Uva, Bill Greene, Bruno Ventelou, Michael Gerfin, Kaspar Wüthrich, Eva Deuchert, Lukas Kauer, David Roodman, Mujaheed Shaikh and seminar participants at the iHEA World congress 2015 in Milan, the 2nd EuHEA student-supervisor conference 2015 in Paris, the 2014 SSPH+ Doctoral Workshop in Health Economics and Public Health in Lugano and the Applied Statistics Seminar at the University of Lucerne for helpful suggestions and comments.

Funding for this project was provided by the Swiss National Science Foundation (Grant 100014/134526).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Human and animal rights The article does not contain any studies with human participants or animals performed by any of the authors.

Appendix

A.1 Variable construction

The indicator for DI benefits is constructed from a question about DI benefit receipt in the past year and equals one for individuals who report to receive DI benefits and equals zero for those who report that they do not receive benefits. There is also information about the amount of DI pensions in the Swiss Household Panel, but we decided against using this variable because of many missing values and the likely presence of reporting error, which would bias the results. Moreover, our instrument likely has a stronger effect on the extensive margin of DI benefits than the intensive margin. The discrete labor supply variable takes three values for individuals who work part-time (1), full-time (2) or are unemployed or not in the labor force (3).

Regarding the control variables, we create a binary variable from the life satisfaction variable (originally scaled from “not satisfied” (0) to “completely satisfied” (10)), which is coded such that it is one for individuals with a satisfaction score of at least seven (the median value) and zero else. General health status ranges from “very well” (1) to “not well at all” (5) and is recoded as a dummy variable good health that takes the value one for individuals with a “very well” or “well” health status, and zero else. Physical activity is a binary variable indicating whether a person exercises for at least half an hour a week (1) or if that person remains inactive (0). Health impediments in everyday activities and medication needed in everyday functioning are measured on an 11-point scale from “not at all” (0) to “a great deal” (10). For both of these variables, we generated an indicator equal to one for individuals with a value of at least 5 in terms of severity of health impediments and medication needed for everyday functioning, and zero for all others. Depression, anxiety and blues are measured on a scale of “never” (0) to “always” (10). We constructed an indicator depression from this information that equals zero for all observations below 3, and one for the rest. The cutoff value of 3 is chosen here as the 75% quantile in the distribution of the original depression variable. Finally, indicators for back problems, weakness and weariness, sleeping problems and headaches are dummy variables, which were created in the way that they are one for observations that report that they are suffering “very much,” and zero for those who are suffering “not at all” or “somewhat” on the original three-point scale.

A.2 Simulated log-likelihood

The vector of model parameters $\theta_j \equiv \left\{\beta_j, \gamma_j\right\}$ and Ω are estimated by maximizing a simulated log-likelihood function of the form,

$$\begin{aligned}\mathrm{SLL}(\theta, \Omega; x, y) = &\sum_{j=1}^{3}\sum_{n=1}^{N} d_{ij} D_i \log(\tilde{P}(y_i = j|D_i = 1)) \\ &+ \sum_{j=1}^{3}\sum_{n=1}^{N} d_{ij}(1 - D_i)\log(\tilde{P}(y_i = j|D_i = 0))\end{aligned} \tag{6}$$

where d_{ij} is an indicator for the choice taken by individual i, D_i is the indicator for the DI benefits and $\tilde{P}(\cdot)$ is the simulated (conditional) choice probability.

Table 8 shows the discrete ES estimates for 1, 10, 100 and 1000 draws per observation, the number of iterations that were needed to reach convergence, the value of the pseudo-log-likelihood at the coefficient vector and the computation time in seconds. All simulations were ran on an Intel(R) Core(TM) i7-4790 CPU @ 3.60 GHz with 16GB RAM on Windows 10 using Stata/MP 14.0. Recall that the main results are based on 1000 draws per observation so that Table 8, column 4 is identical to the results presented in Sect. 6. For small numbers of draws, we know from asymptotic theory that the MSL estimator is not equivalent to the ML estimator and inconsistent (Gouriéroux and Monfort 1991). This is likely the case for the estimates corresponding to $S = 1$ and $S = 10$ draws per observation. For such small S, the reduced computation time comes at the cost of inconsistent parameter estimates. However, as the number of draws is increased, the MSL estimates stabilize and remain basically unchanged. Table 8 suggests that estimates with as few as 100 random draws per observation produce reliable coefficient estimates that are close to those with $S = 1000$. This is also relevant to know from a practical point of view since the differences in computation time are considerable: For the full specification, the computation time is roughly 7 minutes for $S = 100$, but already 18 minutes for $S = 1000$. If the number of draws is further increased,[15] the coefficient estimates hardly change but again come at the cost of a significantly higher computational time.

[15] The specifications were also estimated for $S = 2000$ and $S = 5000$, results available upon request.

Table 8 Sensitivity checks on the number of simulation draws

Number of draws	$S = 1$		$S = 10$		$S = 100$		$S = 1000$	
Working full-time								
DI benefits	− 1.93***	−1.52**	− 2.09***	− 2.13*	− 2.03***	−2.09*	− 2.05***	− 2.04*
	(0.48)	(0.63)	(0.71)	(1.27)	(0.59)	(1.23)	(0.57)	(1.21)
Out of labor force								
DI benefits	0.07	− 0.44	− 0.07	− 0.19	− 0.36	− 0.06	− 0.64	− 0.04
	(0.38)	(0.38)	(0.46)	(1.39)	(0.64)	(1.09)	(0.86)	(1.17)
Background 1: Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Background 2: SES and health	No	Yes	No	Yes	No	Yes	No	Yes
Number of iterations	6	9	7	10	7	9	9	9
Pseudo-log-likelihood	− 2204.381	− 1793.982	−2204.127	− 1794.592	− 2204.015	− 1794.747	− 2203.98	− 1794.72
Computation time (s)	2.301	3.332	8.213	10.207	310.989	451.65	966.41	1106.55

Discrete ES coefficient estimates using different numbers of simulation draws (S). The simulations were ran on an Intel(R) Core(TM) i7-4790 CPU @ 3.60 GHz with 16GB RAM on Windows 10 using Stata/MP 14.0. See the notes of Table 6 for a description of the background characteristics used in each specification

Standard errors clustered at the household level: ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

References

Angrist JD, Pischke J-S (2009) Mostly harmless econometrics: an empiricist's companion. Princeton University Press, Princeton
Autor DH, Duggan MG (2003) The rise in disability rolls and the decline in unemployment. Q J Econ 118(1):157–205
Bound J (1989) The health and earnings of rejected disability insurance applicants. Am Econ Rev 79(3):482–503
Bound J, Burkhauser RV (1999) Economic analysis of transfer programs targeted on people with disabilities. In: Ashenfelter O, Card D (eds) Handbook of labor economics, vol 3C. North-Holland, Amsterdam
Burkhauser RV, Daly MC, McVicar D, Wilkins R (2013) Disability benefit growth and disability reform in the U.S.: lessons from other OECD nations. Working paper series, Federal Reserve Bank of San Francisco
Bunch D (1991) Estimability in the multinomial probit model. Transp Res B 25:1–12
Bütler M, Deuchert E, Lechner M, Staubli S, Thiemann P (2015) Financial work incentives for disability benefit recipients: lessons from a randomised field experiment. IZA J Labor Policy 4:18
Chen S, Van der Klaauw W (2008) The work disincentive effects of the disability insurance program in the 1990s. J Econom 142:757–784
De Jong P, Lindeboom M, Van der Klaauw B (2011) Screening disability insurance applications. J Eur Econ Assoc 9(1):106–129
Eugster B, Deuchert E (2015) Crawling up the cash cliff? Behavioral responses to a disability insurance reform
Federal Social Insurance Office (FSIO) (2013a) Merkblätter, 4.01 - Leistungen der IV. http://www.ahv-iv.info/andere/00134/00186/index.html?lang=de. Accessed 7 Nov 2018
Federal Social Insurance Office (FSIO) (2013b) Merkblätter, 4.06 - Das IV-Verfahren. http://www.ahv-iv.info/andere/00134/00186/index.html?lang=de. Accessed 7 Nov 2018
Federal Social Insurance Office (FSIO) (2014) Statistiken zur Sozialen Sicherheit, IV-Statistik 2013. http://www.bsv.admin.ch/dokumentation/zahlen/00095/00442/. Accessed 7 Nov 2018
Federal Social Insurance Office (FSIO) (2016) Statistiken zur Sozialen Sicherheit, IV-Statistik 2015. https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/publikationen.assetdetail.2722807.html. Accessed 7 Nov 2018
French E, Song J (2014) The effect of disability insurance receipt on labor supply. Am Econ J Econ Policy 6(2):291–337
Frutos EML, Vall Castello J (2015) Equal health, equal work? The role of disability benefits in employment after controlling for health status. Eur J Health Econ 16(3):329–340
Geweke J (1991) Efficient simulation from multivariate normal and student-t distributions subject to linear constraints. In: Keramidas EM (ed) Computer science and statistics: proceedings of the twenty-third symposium on the interface. Interface Foundation of North America, Inc., Fairfax, pp 571–578
Gouriéroux C, Monfort A (1991) Simulation based inference in models with heterogeneity. Ann Econ Stat 20(21):69–107
Gruber J (2000) Disability insurance benefits and labor supply. J Polit Econ 108(6):1162–1183
Gruber J, Kubik JD (1997) Disability insurance rejection rates and the labor supply of older workers. J Public Econ 64:1–23
Hajivassiliou V, McFadden D, Ruud P (1994) Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results. J Econom 72:85–134
Hajivassiliou V, McFadden D (1998) The method of simulated scores for the estimation of LDV models. Econometrica 66:863–896
Kauer L (2014) The effect of cutting disability insurance benefits on labor supply in households. Department of Economics, University of St. Gallen, St. Gallen
Keane M (1990) Four essays in empirical macro and labor economics. PhD Thesis, Brown University
Keane M (1994) A computationally practical simulation estimator for panel data. Econometrica 62:95–116
Kostol AR, Mogstad M (2014) How financial incentives induce disability insurance recipients to return to work. Am Econ Rev 104(2):624–55
Lee DS, Lemieux T (2010) Regression discontinuity designs in economics. J Econ Lit 48:281–355
Maestas N, Mullen KJ, Strand A (2013) Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt. Am Econ Rev 103(5):1797–1829

Marie O, Vall Castello J (2012) Measuring the (income) effect of disability insurance generosity on labour market participation. J Public Econ 96(1):198–210
Marschak J (1960) Binary choice constraints on random utility indications. In: Arrow K (ed) Stanford symposium on mathematical methods in the social sciences. Stanford University Press, Palo Alto, pp 312–329
Miranda A, Rabe-Hesketh S (2006) Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. Stata J 6(3):285–308
Moore TJ (2015) The employment effects of terminating disability benefits. J Public Econ 124:30–43
OECD (2009) OECD employment outlook, pathways onto (and off) disability benefits: assessing the role of policy and individual circumstances. OECD Publishing, Paris
OECD (2010) Sickness, disability and work: breaking the barriers. A synthesis of findings across OECD countries. OECD Publishing, Paris
Parsons DO (1980a) The decline in male labor force participation. J Political Econ 88(1):117–134
Parsons DO (1980b) Racial trends in male labor force participation. Am Econ Rev 70(5):911–920
Roodman D (2011) Fitting fully observed recursive mixed-process models with CMP. Stata J 11(2):159–206
Rothenbühler M, Voorpostel M (2016) Attrition in the Swiss household panel: are vulnerable groups more affected than others? In: Oris M, Roberts C, Joye D, Ernst Stähli M (eds) Surveying human vulnerabilities across the life course: life course research and social policies, vol 3. Springer, Cham, pp 223–244
Shu P (2015) Asset accumulation and labor force participation of disability insurance applicants. J Public Econ 129:26–40
Staubli S (2011) The impact of stricter criteria for disability insurance on labor force participation. J Public Econ 95(9–10):1223–1235
Stock JH, Yogo M (2005) Testing for weak instruments in linear IV regression. NBER technical working paper no. 284, National Bureau of Economic Research
Train EK (2009) Discrete choice methods with simulation, vol 2e. Cambridge University Press, Cambridge