I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15386

How Does Job Coaching Help Disability Insurance Recipients Work While on Claim?

Sébastien Fontenay
Ilan Tojerow

JUNE 2022

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15386

How Does Job Coaching Help Disability Insurance Recipients Work While on Claim?

Sébastien Fontenay
Université Libre de Bruxelles, SBS-EM, DULBEA, CEBRIG

Ilan Tojerow
Université Libre de Bruxelles, SBS-EM, DULBEA, CEB and IZA

JUNE 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

How Does Job Coaching Help Disability Insurance Recipients Work While on Claim?*

We evaluate the effects of a Supported Employment (SE) program aimed at Disability Insurance (DI) recipients with mental conditions. The program is characterized by a "work-first" approach, which includes intensive job coaching and follow-along support. Using a Randomized Control Trial with more than 660 participants over a follow-up period of 18 months, we compare the benefits of this newly introduced program to regular vocational rehabilitation services traditionally used in Belgium. We find that SE increases the probability that DI recipients with mental conditions work while on claim and reduces their reliance on DI benefits. Specifically, we estimate that 18 months after the start of their return-to-work program, participants in the SE group are 9.5 percentage points more likely to be working, and receive 6% less in DI benefits than those in the control group.

JEL Classification: H55, I38, J24

Keywords: employment support, disability insurance, mental health, randomized experiment

Corresponding author:
Ilan Tojerow
Université Libre de Bruxelles
Avenue F.D. Roosevelt, 50
1050 Brussels
Belgium
E-mail: ilan.tojerow@ulb.be

* This study is listed on the AEA registry (AEARCTR-0003378). This research has benefited from the support of INAMI/RIZIV, including Clara-Maria Arbesu Gonzalez, Ellen Wyffels, Fabienne Dubois, François Perl, Frank Seynaeve, Isabelle Fransolet, Jean Alves, Kirsten Van Kelst, Marc Casteleyn, Ronny Meert, Saskia Decuman and Thomas Otte. We also thank Actiris, Article 23 asbl, Forem, ISPPC and VDAB who provide support to DI recipients in the field. A special thank also to Natasia Hamarat and Vanessa De Greef. We thank Basile Dabin for excellent research assistance. Finally, we are grateful to Justine Bondoux and William Parienté for very helpful comments and suggestions, as well as seminar participants at Universitat Pompeu Fabra, 43e Journées des Economistes de la Santé Français, CEBRIG Université Libre de Bruxelles, Royal Economic Society 2022, SOLE 2022, 21st Journées Louis-André Gérard-Varet.

Introduction

In recent decades, the number of individuals receiving Disability Insurance (DI) benefits has increased substantially in OECD countries, creating an important challenge for social security funding (OECD, 2010). In Belgium, where this field experiment takes place, the share of working-age population receiving DI benefits from the long-term program reached 6% in 2017.[1] Among the causes associated with this trend, we observe a rise in beneficiaries with mental conditions (OECD, 2010). As argued by Autor and Duggan (2006), mental disorders are characterized by early onset and low mortality, which increase the average duration of disability spells and hence the size of the recipient population. As a result, the fraction of DI beneficiaries diagnosed with mental disorders has increased over time to represent 35% in Belgium (Gerritse, Marcato, Plasman, & Tojerow, 2017).[2]

The rise in mental disorders has raised questions about what the boundaries are between individuals deemed totally and permanently disabled and others who retain some work capacity or can recover it in the future. Indeed, mental illnesses are often characterized by changing productivity levels over time (Kessler et al., 2006, 2008). For this reason, Bound and Burkhauser (1999) believe that since mental health is more difficult to monitor, individuals with mental health conditions would be on average "healthier" and would have worked in the absence of DI schemes. Previous studies have supported this claim, showing that DI beneficiaries have substantial remaining capacity to work (Bound, 1989; French & Song, 2014; Maestas, Mullen, & Strand, 2013). Maestas, Mullen, and Strand (2013) use examiner assignment as an instrumental variable to show that the employment rate of DI recipients would have been 28 percentage points higher had they not received benefits, with an effect that reaches 37 percentage point for applicants with mental disorders.[3]

[1] In comparison, the share of working-age population in the United-States receiving DI benefits from the federal program increased from 2.2% in the late 1970s to 4.6% in 2013 (Liebman, 2015).

[2] In comparison, the share of DI recipients with mental disorders was 25.4% in the United States (Autor & Duggan, 2006).

[3] French and Song (2014) find that DI benefit recipients reduced labor force participation rates by 26 percent in the United States, but that the labor supply of individuals with mental health conditions was less sensitive to benefit recipients (20 percentage points).

The fact that a substantial fraction of DI beneficiaries retains some capacity to carry on economic activities motivates the implementation of active labor market programs to help them return to work. In this study, we evaluate the effects of a new Supported Employment (SE) program aimed at DI beneficiaries with mental conditions. The program is characterized by a "work-first" approach that includes intensive job coaching. To this end, caseworkers have a maximum of 20 beneficiaries at the same time, which allows them to hold frequent meetings (at least every two weeks). The SE program also offers follow-along support after a job has been found, with the aim of helping beneficiaries secure long-term employment. Using a Randomized Control Trial, we compare the effects of this newly introduced program with regular Vocational Rehabilitation (VR), used for more than a decade in Belgium (i.e., control condition). Regular rehabilitation services favor "human capital acquisition" through vocational training before attempting a return to the labor market. In addition, regular VR offers support that is limited in both duration and intensity, with caseworkers providing services to about 100 beneficiaries. Table 1 summarizes how the two return-to-work programs differ in focus, intensity, and duration.

Between March 2018 and December 2019, we recruited more than 660 DI beneficiaries from across Belgium who suffered from mental illnesses and were willing to take part in a return-to-work program. Participants were randomly assigned to the new SE program or the regular VR program. We followed them for 18 months, from the start of their program, using data from administrative registers and survey instruments designed for this research.

We find that DI recipients with mental conditions who were randomly allocated to the new SE program are twice as likely to work while on claim compared to those in regular rehabilitation. Specifically, we estimate that 18 months after the start of their return-to-work program, participants in the SE group are 9.5 percentage points more likely to be working part-time than those in the control group. Using survey responses, we observe that they find occupations mostly in the private sector (and not in sheltered workshops). At the same time, increased employment in the SE group translates into a reduction in

benefits, paid by the DI, of 6%. In addition, we find that SE was effective in helping DI beneficiaries find and retain a job through the economic turmoil that resulted from the COVID-19 pandemic.

Our rich survey instrument, administered every six months, allows us to explore the channels through which SE achieves a higher re-employment rate than regular rehabilitation, up to 18 months after the start of the program. We find that participants in SE are more likely to be actively looking for a job and dedicate more time to their search (about 2.5 hours per week at follow-up 12) and less time to vocational training (two weeks less, that is, a reduction of about 50% compared to the control group). In addition, DI recipients in SE report that they seek a job with fewer working hours and lower pay (about 10% lower than those in regular rehabilitation). We believe that these differences reflect the fact that SE participants have formed more realistic expectations about their true ability to work, which might only be part-time, while those in regular rehabilitation might seek to fully reintegrate into the labor market. Finally, our survey reveals that SE does not seem to have unintended consequences for the health, perceived self-esteem, or self-efficacy of program participants.

We also take advantage of detailed information provided by the Social Security Administration on their expenditures on both programs in order to perform a cost-benefit analysis. This is particularly important given that the intense supervision provided in the SE program also means higher cost per participant. When taking into account the benefits for the Social Security Administration (in terms of budget saving from the reduction in DI benefits), our most conservative estimates show that the gap between the cost of the two programs would be closed in less than two years. While there is no guarantee that the estimated effects will last beyond the 18-month follow-up period, a dynamic analysis on a monthly basis does not suggest that participants in the control group catch up in terms of their employment rate between follow-ups 12 and 18.

There are only a handful of studies on the effects of programs aimed at helping DI recipients return to work (Broadway & Kassenboehmer, 2019; Fogelgren, Ornstein, Rödin,

& Skogman Thoursie, 2021; Markussen & Røed, 2014).[4] A study by Markussen and Røed (2014) for Norway suggests that subsidized employment on the regular labor market improves the long-term perspectives of temporary DI recipients in comparison to sheltered employment or vocational training.[5] They also show that ordinary education (i.e., in schools, colleges, or universities) has a positive impact on re-employment probability, despite a long lock-in period. In contrast, Broadway and Kassenboehmer (2019) find that job coaching requirements for young DI recipients below the age of 35 in Australia had no effect on the probability of working in the long-run.

Likely the closest paper to ours, Fogelgren *et al.* (2021) use a randomized experiment in Sweden to evaluate the effects of an SE program. They find that the program outperforms existing rehabilitation strategies in terms of rate of return to employment. Their control condition is, however, different from ours, since regular vocational rehabilitation in Sweden is also based on a "work-first" approach but with less intense supervision.[6] In addition, their study focuses on young adults between 19 and 29 years of age, whereas our sample includes DI recipients of all ages. Another unique feature of our study is that in addition to tracking individuals' administrative status and benefits using social security registers, we designed a survey with detailed questions about job search efforts, type of employment,

[4]A study by Laun and Skogman Thoursie (2014) in Sweden examines whether privatization of vocational rehabilitation can improve labor market opportunities for individuals taking a long-term sickness absence. They find no employment differences between private and public rehabilitation providers. Another study by Dean, Pepper, Schmidt, and Stern (2017) evaluates the effect of vocational rehabilitation for individuals with mental illness, but not necessarily on DI rolls. Using a structural model and data for the State of Virginia, they find that VR services have positive effects on participants' future earnings, but in part due to the increased probability of receiving DI benefits. An additional study worth mentioning is the Mental Health Treatment Study (MHTS) conducted by the Social Security Administration (SSA) in the United States between 2006 and 2010. Using a large-scale randomized experiment with more than 2,000 Social Security Disability Insurance (SSDI) beneficiaries who suffered from severe psychiatric impairment (mostly schizophrenia), the SSA measured the effects of access to supported employment services, coupled with systematic medication management. The scope of the treatment was therefore much larger than vocational rehabilitation and included a significant medical component with fully reimbursed mental health treatments. Frey *et al.* (2011) show that compared to the control group, which received no support, SSDI beneficiaries in the treated group had a higher employment rate over the 24-month follow-up period (although with earnings below the SSA's defined "substantial gainful activity" limit, implying that they would stay on DI rolls), as well as improving mental health and quality of life.

[5]The authors argue that sheltered employment or vocational training courses that target DI recipients carry a stigma and send a negative signal to potential employers (Markussen & Røed, 2014).

[6]Fogelgren *et al.* (2021) also use a second control group based on "case management," which is a high intensity support program, but whose main goal is to increase the well-being and social integration of individuals with severe mental conditions, without a clear focus on labor market participation.

vocational training attendance, as well as health, self-esteem, and self-efficacy indices. As such, we are able to paint a much richer picture of program participants' trajectories and explore potential channels for the observed differences in re-employment success. Finally, because we recruited participants up to December 2019, we can also evaluate the effectiveness of SE during the economic turmoil that resulted from COVID-19.

Our study relates to the broader literature on active labor market policies aimed at the unemployed. In a recent review of over 200 studies, Card, Kluve, and Weber (2018) show that the time profile of impacts varies by type of program. They find that job search assistance programs that emphasize "work first" tend to have relatively stable impacts over time, whereas programs that emphasize "human capital accumulation" through vocational training have larger average effects in the medium and longer runs. They also highlight systematic heterogeneity across groups, with long-term unemployed participants benefiting more from "human capital" programs, while "work-first" programs tend to be more successful for participants with low income and/or low labor market attachment.

The remainder of the paper is organized as follows. Section 1 provides more details on the institutional context and the two return-to-work programs evaluated in this research. Section 2 introduces the experimental framework and data collection process. Section 3 presents the results. Section 4 offers robustness checks. Section 5 compares the costs and benefits. Section 6 concludes this study.

1 Institutional Context and Interventions

In this section, we describe the Belgian Disability Insurance (DI) system, focusing on financial conditions. Furthermore, we discuss the differences between the regular Vocational Rehabilitation (VR) program, which has been in place for more than a decade, and the new Supported Employment (SE) program evaluated in this article.

1.1 The Belgian Disability Insurance System

In Belgium, employed workers with a minimum number of working days have access to disability benefits through the National Institute for Health and Disability Insurance (NIHDI).[7] The benefits cover them against health-related events that affect their ability to work for at least one month.[8] Application terms and conditions vary, however, between disability spells that are less than a year and those that are longer. In the rest of this paper, we will therefore distinguish between these two types by referring to "short-term disability" spells and "long-term disability" spells.

To qualify for short-term disability coverage, individuals must be recognized as "unable to work" by a doctor designated by their health insurance fund. Workers would be considered eligible when their ability to work is reduced by at least 66% with respect to their last occupation. To qualify, applicants should also have stopped all productive activity as a consequence of a deterioration in their health that is not directly related to their professional activity.[9] If these two conditions are still applicable after a year, a disabled worker may qualify for long-term disability status. There is, however, no automatic transition from the short-term status to the long-term one. In order to be accepted into the long-term disability program, the applicant's doctor (who oversaw the applicant during the short-term period) must submit the application to the NIHDI, which can directly approve the doctor's decision or run its own internal evaluation.

The replacement rate also varies according to the duration of the disability spell. In the first year, it amounts to 60% of the last wage payment received before becoming disabled. After one year, when one enters the long-term disability program, the replacement rate depends on the last wage payment received as well as the position of the disabled person in the household. To be precise, this share is 65% for heads of households, 60% for single

[7]Full-time workers and unemployed workers must have fulfilled a minimum of 180 working days (or active days of job search for the unemployed) during the last twelve months to be eligible. For part-time workers, the requirement is to have worked at least 800 hours in the last 12 months.

[8]Spells shorter than a month are fully paid by employers and are not covered by this insurance program.

[9]This is to establish a distinction between the disability insurance program and other programs such as the occupational injuries fund and the occupational diseases fund.

households, and 40% for cohabitants, with defined floor and ceiling amounts.[10] Benefits can be reduced when DI recipients start working while on claim. As shown in Figure 1, the benefits remain unchanged if the ratio between hours worked and full-time employment (38 hours a week in Belgium) is lower than 20%. Above 7.6 hours per week, benefits are reduced by the percentage that exceeds the threshold of 20% (for 19 hours per week, this would be 50%-20%, that is, 30%). If, after attempting to go back to the labor market, the beneficiaries become sick again within three months (14 days for short-term program), they automatically requalify for benefits and their previous disability spell is continued.

1.2 Regular Vocational Rehabilitation Program

We start by describing the regular Vocational Rehabilitation program, in place since 2009, which constitutes the control condition to which the new program based on Supported Employment is compared. Before this study, it was the only return-to-work program offered to DI recipients who were seeking help to re-enter the labor market.

Following many other countries, Belgium has traditionally favored vocational training to help DI beneficiaries return to work. This approach relies on the assumption that DI recipients should first rebuild working capacity before they re-enter the labor market. In its current form, the VR program includes three phases supervised by the beneficiary's doctor and a caseworker: 1) orientation; 2) training; and 3) job search assistance. This approach favors human capital acquisition before attempting to return to the labor market.

During the *Orientation* phase, meetings are organized between caseworkers and DI beneficiaries at their local public employment service.[11] The goal of those meetings is to assess rehabilitation needs and find adequate vocational training. The next phase

[10] In 2020, maximum short-term disability benefits were 2,248 euros per month, while maximum long-term disability benefits were 2,435 euros per month.

[11] Job centers in Belgium are a regional competency, supervised in Flanders by the *VDAB*, in Brussels by *Actiris*, and in Wallonia by the *Forem*.

begins when the doctor and NIHDI[12] have approved the *Training* program. Upon training completion, participants receive *Job search assistance* for a maximum period of six months.

From administrative records provided by the Social Security Administration, we know that the median duration of the training is 6 months (first quartile is 3 months, third quartile is 11 months). Respondents to our surveys provided similar information, with median training duration equals to 30 weeks (first quartile is 10 weeks, third quartile is 52 weeks). Our questionnaires also provide additional information on the title and content of the training. Text analysis on the information provided by survey respondents reveals that some followed generic training for computer, accounting or administrative skills, as well as language courses (English, Dutch or French). Others followed specific training to access professions that require basic skills such as "medical secretary or assistant," "forklift operator," "esthetician or beauty therapist," while others have started advanced courses on "design of web and mobile applications," "3D modeling or infographic." The variety of training explains why some last a few weeks, while others take up to a year to complete.

1.3 New Supported Employment Program

We now turn to describing the new Supported Employment program, which is evaluated in this research. It is, in the words of Card *et al.* (2018), a "work-first" program. Indeed, the emphasis is on a rapid, intensive job search so that participants can have face-to-face contact with potential employers from the very beginning of the process. To this end, caseworkers (called "job coaches") have a maximum of 20 individuals listed at the same time and organize meetings at least every two weeks. Vocational training can be offered within the program framework, ideally in combination with part-time employment, or at least after the beneficiary attempted to find work.

The program builds on the IPS model of Supported Employment that was developed

[12]NIHDI must approve training for it to be financed

in the United States in the 1990s.[13] In addition to the rapid job search, caseworkers offer follow-along support, even after a job has been found. The goal is to help beneficiaries secure long-term employment, even though mental illnesses are often characterized by ups and downs that can affect their ability to work. The caseworkers also provide financial advice on the consequences of working for disability benefits. They are also involved with employers at all stages: to find vacancies, make appointment for their beneficiaries, request feedback after job interviews, but also to provide support once the beneficiary started working. As such, they serve as a back-up during unexpected health crises and may reduce uncertainty for employers.

Table 1 summarizes the main aspects of the Supported Employment program and how it differs from the regular rehabilitation approach. One may notice that the caseworkers' load is about five times smaller for the new Supported Employment program. In addition, the program favors competitive work on the regular labor market and not sheltered or wage-subsidized employment.

2 Experimental Design and Data Collection

In this section, we present the experimental setup used to measure the effectiveness of the SE program in comparison to regular VR. We also describe the outcomes constructed to compare the effects of the two programs using both administrative registers and a survey designed for this research project.

[13]The Individual Placement and Support (IPS) approach was developed primarily to provide Supported Employment services for people with severe mental health conditions (e.g., schizophrenia, bipolar disorders) at mental health centers. The program was first evaluated by Drake, McHugo, Becker, Anthony, and Clark (1996) at two mental health centers in New Hampshire. Many small-scale randomized controlled trials followed (summarized in Luciano et al., 2014 and Marshall et al., 2014). A recent meta-analysis by Modini *et al.* (2016) reveals that the IPS model was effective in contexts characterized by heterogeneous labor market conditions.

2.1 Recruitment of Participants and Random Allocation to Programs

In March 2018, NIHDI notified all healthcare-funded doctors in Belgium that a study was underway to evaluate the effects of a new return-to-work program for DI recipients with mental conditions. The doctors, in turn, informed patients who (1) suffered from mental illness and (2) were willing to re-enter the labor market that they were eligible to participate in the study.[14] Participants were recruited in Belgium's three regions. Wallonia was split to account for two different partners assuming responsibility for the implementation of the Supported Employment program there.[15] In total, the experiment comprises four clusters that cover all of Belgium (i.e., Flanders, Western Wallonia, Eastern Wallonia, and Brussels). Doctors were encouraged to recruit patients with moderate to severe mental disorders. The form sent to doctors explains that moderate disorders may include depression or anxiety, while severe disorders may include bipolar disorders, schizophrenia or obsessive compulsive disorders. The list was not meant to be exhaustive, but doctors were told not to include patients with mild mental health issues who should be able to return to the labor market on their own.

DI recipients who agreed to take part in the study signed an informed consent before their doctor sent their file to NIHDI and the researchers. Their administrative file contained individual characteristics (gender and work experience) used for the stratified randomization, as well as names and contact details for the follow-up surveys. Randomization was performed at the individual level. On a weekly basis, we allocated participants to the treatment or control groups using the randomization list for the correct cluster and stratum.[16] We used stratification on two variables that are important confounding fac-

[14]The doctors were asked not to discuss the specifics of the return-to-work programs so as not to create expectations.

[15]Wallonia was divided between West and East, based on the postal codes of participants. There was thus no overlap between the two sub-regions.

[16]The randomization took place at the individual level and was performed on a weekly basis by the researchers. As participants entered the study over several months between March 2018 and December 2019, a predefined allocation sequence was created in the form of computer-generated randomization lists. The randomization lists were created using the software Stata and the user-written command "ralloc" from Ryan (1998). The command provides a sequence of treatments randomly permuted in blocks of

tors of return-to-work success according to previous studies: gender (Wewiorski & Fabian, 2004) and work experience in previous years[17] (Drake et al., 1996).

Participants were then invited to a job center in their region to complete the baseline survey. After completing the baseline survey, they met with a caseworker who informed them of the program in which they had been randomly allocated.[18] In total, 667 participants were recruited from March 2018 to December 2019.

2.2 Administrative and Survey Data

We rely on two data sources to estimate the effects of the two return-to-work programs: (1) administrative registers and (2) survey instruments designed for this research. The main advantage of using administrative data is that we can observe the complete sample of participants for the duration of the experiment and at a relatively high frequency. At the same time, conducting a dedicated survey allows us to collect detailed information on individual behavior and well-being.

The administrative data comes from the registers of NIHDI. They provided pre-treatment characteristics of participants, including age, gender, place of residence, date of entry into DI, status before entering DI (i.e., a blue-collar or white-collar worker), and whether they have dependents (children or spouse). On a monthly basis, we are able to track: their disability status; benefit amount (in 20-euro bins); participation in vocational training financed by NIHDI; and most important, partial work resumption (with including the exact volume of work). Our administrative dataset allows us to build a balanced panel for our entire sample of participants that spans 12 months before the start of their

varying size (treatments are balanced within blocks). In total, 16 randomization lists were created, one for each combination of the four regional clusters and the two stratification variables (i.e. gender and work experience). Each week the research team at the university received a list of new participants that had signed the informed consent form. New participants were allocated to the treatment or control groups using the randomization list of the correct cluster and stratum.

[17]The exact question is: "Have you been working in a paid job in the last two years?" The answer is binary (yes/no), but an additional third category covers cases where the information was not provided by the doctor.

[18]Caseworkers were asked not to give any information about the other program.

return-to-work program and 18 months after it.

We complement the data from administrative registers with a survey designed for the purposes of this research. All participants answer a baseline survey before the start of their return-to-work program. The baseline survey instrument allows us to collect a rich set of information on all participants, including their nationality, education, previous work experience, and financial situation. All participants also receive a follow-up questionnaire every six months, by email or in hard copy. Follow-up surveys provide us with detailed information on meetings with caseworkers, job search efforts, labor force participation (including the type of contract and sector of activity), and earnings sources, as well as subjective evidence on their health and well-being.

2.3 Integrity of the Experimental Design

Balancing test: Table 2 presents summary statistics for DI beneficiaries before their program assignment (using both administrative and survey data). The first column displays means and standard deviations in parentheses for the entire sample of 667 participants. The next two columns show the respective statistics for the control and treatment groups. The statistics reveal that participants are perfectly balanced in terms of gender. Their average age is 40 years and they have spent 44 months on disability. Only 20% have dependents, either children or a spouse. 53% were blue-collar workers before entering DI, while 26% had a higher education degree. Finally, the vast majority holds Belgian nationality (87%).

The last column in Table 2 reports results from balancing tests. The latter reveal that at the 10% level, we fail to reject the equality of means of treatment and control groups for any of the 11 outcomes considered. The aggregate test, reported in Panel B, also finds that we are not able to reject equality of means across all 11 variables (p-value = 0.65). Overall, it appears that individuals in both groups showed similar characteristics.

Survey attrition: Table 3 presents an analysis of survey attrition for the follow-up sur-

veys at 6, 12, and 18 months. The follow-up rate is relatively good for a self-administered questionnaire, with about 64% of participants who answered the first survey at 6 months. Panel B presents an analysis of the type of people who were less likely to be surveyed. Those who did not answer were slightly more likely to be men of foreign nationality and without higher education. Panel C presents a test of whether the treatment affected the type of person who completed follow-up surveys, in other words, whether the treatment caused a sample composition bias. The p-values on a full set of baseline characteristics interacted with treatment are 0.62 (follow-up 6 months), 0.31 (follow-up 12 months) and 0.97 (follow-up 18 months). Taken together, these results demonstrate that there is no systematic difference between individuals who answered the follow-up survey in the treatment and control groups.

3 Results

In this section, we introduce our empirical strategy and subsequently compare the effects of the two return-to-work programs on the set of outcomes described in the previous section.

3.1 Estimation Strategy

We estimate specifications that compare the effects of the new SE program (treatment) with the regular VR (control). We run the following Ordinary Least Squares regression:

$$Y_i = \alpha + \beta Treatment_i + \gamma X_i + R_{region} + S_{month} + T_{year} + \epsilon_i \quad (1)$$

where Y_i is the relevant outcome for DI beneficiary i, $Treatment_i$ is an indicator variable equal to one for beneficiaries who were randomly assigned to the new Supported Employment program, X_i denotes a vector of beneficiaries controls that includes the two

stratification variables (dummies for gender and work experience in last two years), as well as the individual's baseline value of the outcome variable Y (when available). R_{region} is a vector of dummy variables for the four regions in the study (described in subsection 2.1). For estimations using the administrative register data at monthly frequency, we also include month S_{month} and year T_{year} fixed effects to account for seasonality and trends. The coefficient of interest is β and captures the effect of being assigned to treatment. In all estimations, we report heteroskedasticity robust standard errors.

3.2 Effects on Administrative Status and Benefits Payment

We start our analysis by comparing the effects of the two return-to-work programs on the DI status and benefit receipt of study participants over time. We use data from NIHDI administrative registers, which allows use to track all participants during a 30-month period, including 12 months before and 18 months after the start of their return-to-work program. We create three dummy variables that track the disability status of individual participants on a monthly basis, and whether they work or follow a training while on claim.[19] We also create a variable to report monthly DI benefits received by participants, which takes the value 0 if the individual exited DI. As such, the benefits variable will capture effects on DI reliance at both the extensive and intensive margins.

Figure 2 compares the trajectories of DI recipients who randomly joined one of the two return-to-work programs. One can observe in Panel C that their probability to be on disability rolls is highly similar across the 30-month window. Panel A, however, shows that SE participants are more likely to start working part-time while on claim from the 6^{th} month after the start of their program, with a gap that widens up to the 18^{th} month. At the same time, Panel D reveals that SE participants receive lower benefits in the long-run, which is expected when DI recipients work part-time. Finally, Panel B reveals

[19]The two variables imply contradictory effects for the social security budget. Indeed, as explained in subsection 1.1, individuals who work while on claim will receive reduced benefits if their working hours exceed 20% of a full-time job. In contrast, the cost of training, when approved by the Social Security Administration, is fully covered. In addition, participants receive a bonus for each effective hour of training, as well as a final bonus if successfully the training is successfully completed.

that SE participants are less likely to follow a training while on claim, compared to VR participants whose program focuses on vocational rehabilitation.

Table 4 presents formal estimates for the difference in trajectories between participants in SE and VR. We provide results for the four outcomes described above at months 6, 12, and 18 after the start of a return-to-work program, which match the follow-up periods of the survey and should ease the comparison. Since administrative data are available on a monthly basis, we also produce four graphs that show the evolution of the effects for all outcomes over time (Figures 3, 4, 5 and 6). Figure 3 clearly shows that SE does not seem to increase the probability of exiting entirely from DI compared to the regular rehabilitation program. The estimated intention-to-treat effects are very small and their confidence intervals always include 0. However, Figure 4 suggests that those who took part in the new SE program are more likely to work part-time while on claim. This effect starts around 6 months after program entry and slowly builds, reaching 9.5 percentage points after 18 months (equivalent to two times the control mean). At the same time, participants in the SE program are 2.5 percentage points less likely to follow a training at follow-up 18. This is not surprising given that SE relies on a "work-first" approach, while regular rehabilitation in the control group clearly encourages vocational training. Finally, when looking at the consequences for disability payments, Figure 6 clearly indicates a small reduction in the amount received by individuals in the treatment group. Again, this effect materializes around the 6^{th} month after program entry and slowly increases to reach 70 euros per month at follow-up 18, that is a reduction of 6% compared to the control mean. As explained above, the reduction in DI benefits very likely reflects the fact that individuals working more than 20% of a full-time job automatically incur a reduction in their allowance. Taken together, these results suggest that participation in the SE program does not increase the probability of exiting DI, but rather favors working while on claim. The result is less reliance on DI benefits.

3.3 Effects on Type of Employment and Sector of Activity

In the previous subsection, we showed that SE increases the probability that DI recipients work while on claim. We complement this finding using our survey data. The latter has two additional advantages compared to NIHDI administrative registers: (1) it also tracks the employment status of those who left DI (i.e., individuals who returned to full-time employment); and (2) it contains detailed information on the type of contract and sector of activity.

The first column in Table 5 displays the effect of SE on the probability of working in a paid job at follow-ups 6, 12, and 18 months. When last observed at 18 months, the employment rate of the SE group was more than two times larger, with a treatment effect of 8.7 percentage points compared to a control group mean of 13 percentage points. The effect is relatively aligned with the one measured using administrative registers (9.5 percentage points in Table 4), reinforcing the strength of our findings and the reliability of our survey data.

We now turn to the type of contracts that participants in the study have found in the course of their return-to-work process. We distinguish between three types of contract, permanent (i.e., open-ended contract), temporary contract (i.e., fixed-term contract) or self-employment. We believe that the type of contract reflects the quality of the employment found, as well as the strength of the labor market attachment. Indeed, a potential pitfall of the "work-first" approach of SE is the risk of individuals accepting low quality jobs to quickly re-enter the labor market. Previous research in the context of unemployment insurance reveals that job search assistance does not boost employment in the long-run if the program places participants in lower-quality jobs (Cottier, Flückiger, Kempeneers, & Lalive, 2018). Our survey data allows us to check whether this is the case for our study population of DI beneficiaries.

Table 5 shows that the SE group is 6.4 percentage points more likely to find a permanent contract at follow-up 12 (column (2)), but the effect tends to fade over time,

reaching only 2.6 percentage points at follow-up 18 (although with large standard errors). At the same time, we observe a larger probability that SE participants work with a temporary contract at follow-up 18 (column (3)) and no significant difference in terms of self-employment. These results suggest that SE did favor a quicker return to the labor market, but for a significant share of DI recipients, this is only with temporary contracts.

Finally, we take advantage of our survey data to observe the sector of activity for those who found a job. A key feature of the IPS model of SE is to help individuals with mental health conditions find a job in the "regular" labor market, that is avoiding placement in sheltered work. Table 5 reveals that at follow-up 18, the vast majority of those who work in a paid job have a position in the private sector (column (5)), while there is no significant effect on employment in the public or nonprofit sectors. More important, we do not observe any effect on the probability of joining sheltered workplaces. Taken together, our results suggest that participation in SE increases the employment rate of DI recipients with mental conditions, mostly through occupations in the private sector.

3.4 Effects on Earnings

In this subsection, we explore the effects of participation in the two return-to-work programs on earnings. In our survey, we ask participants about three earnings sources: work income; DI benefits; and Unemployment Insurance (UI) benefits. If a participant does not receive earnings from a given source, the variable takes on the value "0," as such, capturing effects at both the extensive and intensive margins. We also create an outcome that sums up earnings from all three types and title it "total earnings."

Table 6 reveals that 18 months after the start of their return-to-work program, participants in the treated group declared receiving on average 119 euros more from work income than those in the control group (Column (2)), that is, about twice as much as the control group. Meanwhile, their DI benefits decrease by 105 euros (Column (3)), that is,

a reduction of 9.6% compared to the control mean.[20] For two reasons, it is expected that the effects on wage income and DI benefits are not symmetrical. First, as explained in subsection 1.1, DI benefits replace 60% of lost income, which means that if someone exits DI and goes back to full-time employment, their wage income should be higher than their benefits. Second, when DI recipients work while on claim, their DI benefits are reduced, but the first 20% are exempted (more details in sub-section 1.1). For both these reasons, it is expected that DI beneficiaries who return to a full-time job or work while on claim will have higher earnings in total.

3.5 Effects on Health and Well-being

In this subsection, we explore the consequences of the programs for the health and well-being of participants. Our survey instrument designed for this project includes three sets of questions that allow us to build the most common indicators used in the literature in health and psychology. First, we ask twelve questions based on the short-form health survey validated by Ware, Kosinski, and Keller (1996). It includes questions on both physical and mental aspects in order to assess the impact of health on an individual's everyday life. Second, we ask survey participants to answer ten questions related to their self-esteem. These ten questions have been used since the seminal work of Rosenberg (1965) to measure both positive and negative feelings about oneself and to detect self-esteem problems. Third, we build on work by Schwarzer and Jerusalem (1995) and include ten questions to measure self-efficacy. The goal of this index is to observe how survey participants perceive their ability to cope with difficult demands in life.

Even though the primary goals of the return-to-work programs are not to improve participants' health or their self-perceptions, contacts with caseworkers and (potential) employers might have unintended consequences. Table 7 shows that SE does not seem to disproportionately affect the health, perceived self-esteem, or self-efficacy of program

[20]The effect is slightly larger than when measured with the administrative registers (Table 4), which could reflect the fact that survey respondents work on average more hours per week and therefore incur a larger reduction in benefits.

participants. We do not observe any significant effect for all three measures at any point in the follow-up period.

3.6 Suggested Mechanisms: Job Search Behavior and Vocational Training

In the previous subsections, we showed that participants in SE are more likely than those on regular rehabilitation to work while on claim, and as a result, rely less on DI benefits. We can think of at least three reasons why SE achieves higher part-time employment rates for DI recipients: (1) they receive more intense support from caseworkers; (2) they more actively look for a job; and (3) they dedicate less time to vocational training before attempting to go back to the labor market. Our survey will help test whether these hypothesized channels are met in practice. Our questionnaire helps us track the frequency of meetings between participants and their caseworkers, as well as the time that is dedicated to the job search or training.

Column (1) in Table 8 shows that during the first six months of participation in a return-to-work program, DI recipients in SE have had on average 4.8 more meetings with their caseworker compared to those in regular rehabilitation (that is, more than twice as many than in the control group, which had 3.8 meetings). Over time, the number of meetings with caseworkers tends to decrease, but it does so at a faster rate in the control group. Thus, when asked how many times they met with their caseworker at follow-up 18, participants in the treated group responded that they had 3.1 more meetings over the previous six months than their counterpart in the control group (mean of 1.9). These results indicate that SE does indeed offer more intensive guidance to DI recipients who are willing to re-enter the labor market.

Our questionnaire also allows us to track whether DI recipients are actively looking for a job and how much effort they commit to it. They are asked how many average hours they dedicate each week to seeking job offers, working on their CV and cover letters, preparing

and going to interviews with potential employers, and other activities related to the job search (e.g., attending job fairs). We combine answers to these four categories and create a variable with the total number of hours dedicated to the job search. If a participant responds that they do not actively look for a job, the variable takes on the value 0. As such, it captures job search efforts at the extensive and intensive margins. Column (2) in Table 8 reveals that at each follow-up period, SE participants are more likely to declare that they are actively looking for a job. When it comes to how much efforts they put into it, we observe that at follow-up 12, participants in SE dedicate on average 2.5 hours more each week to their job search compared to those in regular rehabilitation (Column (3)). This effect fades away at follow-up 18, which most likely reflects the fact that many participants have already found a job and are no longer actively looking for employment.

We also ask survey respondents what their reservation wage is (net of social security contribution and income tax), as well as their preferred number of working hours when looking for a job. Interestingly, at follow-up 6 / 12, participants in SE reported that they preferred to work on average 2.4 / 3 hours less per week, that is, 10% / 12% lower than their control group counterparts (Column (4) in Table 8). They also declare that they would accept a lower wage of about 143 euros, that is, 10% lower than the control mean (Column (5), follow-up 12 in Table 8). We believe that these results could indicate another potential channel for the higher re-employment rate of SE participants. We hypothesize that individuals taking part in regular rehabilitation, which focuses on vocational training to regain work capacity, anticipate a return full-time employment. In contrast, our results seem to indicate that those in SE have lower expectations in terms of working time and wages, which might allow them to apply to a wider range of jobs.

Finally, our survey offers a way to track training attendance, even for short vocational trainings of only few hours or days. This is a truly positive feature compared to administrative registers, which only report on larger training periods for which NIHDI authorization is required and financial support offered. Column (6) in Table 8 shows that participants in SE have spent on average two weeks less on training at follow-up 12

and 18, this is about 50% less than in the control group. This result clearly reflects the different focus of the two programs, but also the fact that SE participants have more time available to look for a job and are not locked into a training that might delay their return to the labor market.

3.7 Heterogeneity analysis

In this last subsection, we explore the heterogeneous effects of the new SE program across chosen sub-groups of participants. We are particularly interested in the heterogeneity of results along six individual dimensions: location, gender, time on disability, blue/white-collar status, education, as well as the participants' belief in their capacity to overcome difficulties. In what follows, we highlight the reasons why we believe these distinctions are of interest for the research on DI and eventually present the estimated differences between groups. All subgroups are constructed around two mutually exclusive categories that encompass the entire sample of participants. We test for the difference between the two categories by adding an interaction term to equation (1), which becomes:

$$Y_i = \alpha + \beta Treat_i + \delta Categ_i + \zeta Treat_i * Categ_i + \gamma X_i + R_{region} + S_{month} + T_{year} + \epsilon_i \quad (2)$$

Figures 7, 8, 9 and 10 report the estimated difference ζ along six dimensions for the probability of exiting DI, the probability of working part-time, the probability of following a training, and the benefits amount.

Geographical location: We start by testing for differences in the effect of SE based on the region where participants lived in at the start of their return-to-work program. We distinguish between Flanders in the north of Belgium and Wallonia/Brussels in the south/center of the country. We are interested in this distinction because Flanders has a more dynamic job market, and according to Eurostat, an unemployment rate of 3.2% in 2019, compared to 7.2% in Wallonia and 12.6% in Brussels. Results in Figure 8 indeed

show that participants in the new SE program were more likely to work part-time in Flanders, while there is no significant difference with VR in Wallonia/Brussels. We believe that this reflects the fact that it is harder for DI recipients to go back to work, even on a part-time basis, in a weak labor market where they compete with a large share of unemployed people.

Gender: We now look into the heterogeneity of the results by the gender of participants. We know from previous studies that the number of women on DI rolls has increased substantially in recent decades (Autor & Duggan, 2006), and particularly among young women after motherhood (Fontenay & Tojerow, 2020). We are therefore interested in the effects of the new SE program for this growing group of female DI recipients. Our estimates, using a gender dummy interacted with the treatment status, do not show statistically significant differences for the four outcomes considered. Despite large standard errors, however, we notice that men are 7 percentage points more likely to work part-time after entering the SE program than women. A better powered experiment might help reveal possible differences in the impact of SE across genders.

Time on disability: As explained in subsection 1.1, Belgian DI distinguishes between beneficiaries who have spent less than 12 months on DI rolls (short-term program) and those who have spent more (long-term program). We use this distinction to evaluate the effects of SE depending on the duration of the disability. Interestingly, we observe in Figure 9 that long-term beneficiaries who take part in the SE program are less likely than short-term beneficiaries to follow a training while on claim. While it was expected that because of the nature of the program, SE participants would be less likely to attend a training, we anticipated that long-term beneficiaries might need to regain human capital before attempting to return to the labor market. We can think of two reasons for this somewhat surprising result. First, long-term beneficiaries might have more stable mental health conditions, allowing them to more quickly re-enter the labor market without a need for training. Second, long-term beneficiaries have been away from the labor market longer and might be less willing to engage in long training that would further delay their

return.

Blue/white collar, education level: We also explore how the effects of SE vary according to workers status (i.e., blue vs white collar) and their level of education (primary/secondary school vs higher education). The interaction effects with the two different dummies do not reveal statistically significant differences in the probability of exiting DI, the probability of working part-time, the probability of following a training while on claim, and the amount of benefits. We therefore conclude that SE works indifferently for blue-collar and white-collar workers, as well as for individuals who attended higher education or not.

Self-efficacy: Finally, we investigate the effects of the SE program based on the participants' belief, before the start of the return-to-work program, that they would be able to overcome difficulties and obstacles. To do so, we use the Generalized Self-Efficacy Scale (GSES) based on previous work by Schwarzer and Jerusalem (1995). The GSES includes ten questions that we report in table 9. We construct a self-efficacy score (ranging from 10 to 40) for all participants based on their answers to the dedicated in module in the baseline survey, that is before their entry in a return-to-work program. We then compute the z-score for each individual by subtracting from their score the sample mean and dividing by the standard deviation. Finally, we distinguish between those who have a low or high self-efficacy depending on whether they are below or above 0. The survey participants who obtain a higher score on the self-efficacy scale have stronger beliefs in their capacity to perform novel or difficult tasks, as well as to cope with adversity. Interestingly, when interacting the self-efficacy dummy with the treatment variable, we observe that those in the SE group who have a high self-efficacy score are 11.3 percentage point more likely to be working part-time than those who have a low score (Figure 8). We interpret this as a sign of the readiness of the participant to cope with the difficult task of returning to the labor market after a long time on disability. The GSES is an operative construct, which is highly correlated with subsequent behavior (Jones, Mandy, & Partridge, 2009; Schwarzer, 1992). As such, we believe that it could be used by doctors or caseworkers to asses the

readiness of the candidates to participate in a return-to-work program, especially for SE which confronts participants to the labor market without prior training.

3.8 Impact of the COVID-19 Pandemic

Previous research on the consequences of economic shocks for beneficiaries of Active Labor Market Policies (ALMP) shows contrasting results. Barrera-Osorio, Kugler & Silliman (2021) find that the COVID-19 pandemic washed away the benefits of a job-training program in Colombia. Field, Linden, Malamud, Rubenson, & Wang (2019) also reveal that cohorts that graduate from vocational programs during economic downturns perform worse. However, others such as Beuermann, Bottan, Hoffmann, Jackson, & Vera Cossio (2021) find that beneficiaries of ALMP suffered fewer employment losses throughout the COVID-19 pandemic. It therefore remains unclear whether pre-pandemic programs might sustain their benefits throughout the turmoil that resulted from COVID-19.

NIHDI and its partners made sure that both return-to-work programs remained active during the pandemic and that caseworkers continued to meet participants virtually. This is not surprising given that Belgium is the country in the European Union with the highest share of teleworkers (above 50%) and a below average share of unemployed since the onset of the pandemic (Eurofound, 2020). For all these reasons, we believe that the impact of the COVID-19 pandemic was more moderate in Belgium, although obviously not null.

In a first exercise, we compare participants who entered the study early, from March 2018 to December 2018, with participants who entered at a later stage between January 2019 and December 2019. The rationale for this comparison is that the follow-up period of late entrants overlaps with the development of the COVID-19 pandemic, which started in March 2020 in Belgium. We plot in Figure 11 their probability to work part-time depending on whether they are early entrants (Mar.-Dec. 2018, left panel) or late entrants (Jan.-Dec. 2019, right panel). We notice a downward shift in the probability of working while on claim for late entrants ("2019 sample") in both groups. The gap

between treatment and control participants at follow-up 18 is 12 percentage points for the "2018 sample" and declines to 8.2 percentage points for the "2019 sample." However, the relative gap compared to the control mean[21] is rather stable over time from 90% to 80%. This graphical exploration seems to suggest that even though late entrants in the study seem to have been affected by the COVID-19 pandemic, those who were randomly allocated to the SE program were still more likely to work while on claim than those in the VR program.

We now turn to more formal estimates of the effect of the COVID-19 pandemic. Since new participants entered the study continuously between March 2018 and December 2019, they have been impacted at different stages of their return-to-work program. For instance, those who entered in December 2019 had only three months pre-pandemic, while those who entered in January 2019 were hit toward the end of their 18-month follow-up. We leverage this unique feature of our study to estimate the impact of the COVID-19 pandemic on the trajectory of our study participants. We estimate the following model to track the effect of COVID-19 over time:

$$Y_{it} = \alpha + \beta Treat_i * Followup_t + \delta Treat_i * Followup_t * COVID_t + \gamma X_i + R_{region} + \epsilon_{it} \quad (3)$$

where the $Treat_i$ indicator is now interacted with a $Followup$ variable that tracks the number of months since the start of the return-to-work programs from zero to 18. The third term of equation 3 adds an interaction with a binary indicator $COVID$ which takes on a value 1 after March 2020. Thus, the vector of coefficients β captures the effects of being assigned to treatment over time before the pandemic, while the vector of coefficients δ captures the effects of the pandemic on the treated. Compared to previous estimations, we now pull together data for all participants during all follow-up periods (i.e., panel dataset). We also cluster standard errors are at the individual level.

Table 10 tracks the effects of the SE program on the probability of working while

[21] The control mean at follow-up 18 is 13.2 percentage points for the "2018 sample" and 10.4 percentage points for the "2019 sample."

on claim from follow-up 6 to 18, and allows comparison to pre-pandemic impacts (first column) and impacts after COVID-19 hit Belgium in March 2020 (third column). We observe in column (1) that the probability of working part-time for participants in SE was already significantly higher at follow-up seven and increased over time, reaching 14 percentage points at follow-up 18. Column (2) reveals that the COVID-19 pandemic drastically reduced the positive effects for the treated in the medium-run (particularly between follow-ups nine and 13). However, we notice in column (3) that in the long-run, participants in SE whose follow-up overlapped with the COVID-19 pandemic were still 10.6 percentage points more likely to be working while on claim compared to those in regular rehabilitation. In other words, the COVID-19 pandemic delayed the return-to-work process of SE participants, but in the long-run, the effects of the program are still largely positive. We conclude that SE was effective in helping DI beneficiaries find and retain a job through the economic turmoil that resulted from COVID-19.

4 Robustness checks

This section offers two robustness checks that have become standard in the literature using randomized controlled trials (e.g. Cohen & Dupas, 2010; Fujiwara & Wantchekon, 2013). We start by showing that, despite the modest size of our sample, the results are very similar when using randomization inference instead of classical inference. In continuation, we account for the fact that we test the effect of the SE program on multiple outcomes and provide p-values accounting for the risk of false discovery.

Randomization Inference was first proposed by Fisher (1935) and further developed by Rosenbaum (2002) as an alternative for classical inference in a randomized experiment context. The main advantage of this procedure is providing inference with correct magnitude regardless of sample size. In addition, this test is nonparametric as it does not make distributional assumptions. Also known as permutation test, this method consists of reassigning the treatment and control status in the sample (in our case within the

strata described in sub-section 2.1) and reestimating the parameter of interest (our β in equation 1) using this placebo assignment multiple times (we perform 1,000 random permutations).[22] This procedure gives p-values for the null hypothesis of zero treatment effect, which corresponds to the proportion of reestimated β that are larger (in absolute value) than the actual β. In Table 11, we report the p-values computed this way under the name "rand. inf. p-value" for all the outcomes considered in our previous estimations. One can see that these newly-computed p-values are very close in magnitude to those from the classical inference method used in our main analysis.

Multiple hypotheses testing: We also want to account for the fact that we are estimating the effects of the new SE program on 25 individual outcomes by adjusting p-values for multiple inference. In particular, we compute sharpened q-values following the procedure by Anderson (2008). This method controls for the false discovery rate, that is the expected proportion of rejections that are type I errors (i.e. false rejections). Table 11 shows the p-values computed this way under the name "sharpened q-values." As expected, the q-values of each test are adjusted upward to reduce the probability of a false rejection. However, one can observe that the effects measured for our main outcomes of interest, those related to the employment status of the participants, remain statistically significant at conventional levels.

We take all these results as a confirmation that, despite the limited size of our sample and the multiplicity of the outcomes tested, the effects measured for the new SE program are genuine and robust.

5 Cost-benefit Analysis

In this section, we perform a cost-benefit analysis of the two interventions to put in perspective the intention-to-treat effects measured previously. Given that the new SE

[22] We use the Stata package "ritest" developed by Simon Heß (2017) to perform the randomization inference procedure.

program requires more intense supervision than regular VR and therefore supposes higher cost per individual participant, this is particularly important. As such, it is not obvious that a program should be preferred over another if its marginal benefits do not make up for its higher costs in a reasonable time frame.[23]

As explained in subsection 1.2, the NIHDI partners with regional employment agencies that offer rehabilitation services to DI recipients. The cost for each program participant is laid out in a cooperation agreement that was renewed in 2018 for the start of this study. NIHDI agrees to pay 4,800 euros per year for each participant in the SE program.[24] Over the course of the study, that is, 18 months, the cost was thus 7,200 euros for each SE participant. In comparison, NIHDI agrees to pay a flat fee for each participant in regular VR of 4,800 euros.[25] The difference between the two interventions was therefore 2,400 euros per individual participant over the 18-month study window. It certainly reflects SE's more intensive investment in human resources than VR. Those costs cover the salaries of caseworkers and their supervisors, as well as all necessary expenditures to perform their mission (e.g., office space, communication devices to stay in touch with program participants, transportation costs to meet prospective employers). It is therefore a comprehensive proxy of the individual cost if the programs were to be scaled up.

When it comes to the benefits of the program, we distinguish between two perspectives: (1) NIHDI budget and (2) the "society as a whole." First, from the perspective of NIHDI, benefits are measured by the reduction in DI benefits paid, which could be the result of individuals exiting DI completely or working while on claim. Second, from the perspective of "society as a whole," the benefits encompass the value of the production generated by new jobs, as well as the savings for NIHDI's budget, which is financed by taxpayers. Following Fogelgren *et al.* (2021), we hypothesize that the production generated by new

[23]See for instance Crépon, Gurgand, Kamionka, and Lequien (2012) for an example of program whose positive effects are small relative to the cost of implementation.

[24]For the ease of calculation, we use the cost per participant in Flanders for the whole country. This is not a strong assumption since Flemish participants make up 70% of our sample. In addition, the cost in Brussels and Wallonia is slightly smaller, so our estimates are in fact the most conservative.

[25]The reason for the flat fee is that, contrary to SE, job coaching in regular rehabilitation is limited to six months after training has been completed. In addition, no support is provided while participants are enrolled in training.

jobs is estimated by the wage cost.[26] This of course assumes that there are no displacement effects and that participants in one program do not crowd out jobs for participants in the other program, or for nonparticipants. Crépon, Duflo, Gurgand, Rathelot, & Zamora (2013) show in the context of unemployment insurance that this type of externality can drastically reduce the estimated benefits of a program, particularly in weak labor markets. In our particular context, we argue that this should not play a significant role since the Belgian labor market was rather strong during the experiment window, especially in Flanders where the unemployment rate was 3.2% in 2019, according to Eurostat. Tensions in the Flemish labor market are one reason that the employment agency was keen to implement activation policies for DI beneficiaries. For all these reasons, we assume that the benefits for the "society as a whole" are the production generated by the new jobs (valued at cost) and the budget savings for NIHDI.

In Table 12, we report the effects of the new SE program on DI benefits and wage income of participants (already shown in Table 6), as well as the combined effect for the "society as a whole," which corresponds to the sum of the absolute values of the two others. However, our sample of survey respondents, while perfectly balanced across programs, is not fully representative of the entire population of DI recipients with mental conditions in Belgium. In fact, in Table 2 we show that participants in the study are more likely to come from the north of the country (i.e., Flanders) and are equally likely to be a man or a woman, as well as a blue or white-collar worker. Population data from NIHDI's registers reveals instead that blue-collar workers and women are over-represented among DI recipients with mental conditions (59% and 61% respectively), while only one out of two beneficiaries comes from Flanders. To improve the representativeness of our study, we reweight our sample using the entropy balancing method by Hainmueller & Xu (2013). The results using this balanced sample are reported under the title "reweighted sample."

At this stage, we calculate how long the benefits of SE would need to last to make up for their higher cost, and we report the corresponding "catch-up time" in Table 12.

[26] Unfortunately, we only know the wage income of employees and not the total wage cost for their employers. This will therefore be a lower bound estimate for the value generated by the job.

From the perspective of the Social Security Administration, the reduction in monthly DI benefits is between 105 euros and 147 euros at follow-up 18, depending on whether we use the survey sample or the reweighted sample. If this effect were to remain stable beyond the 18-month window, the 2,400-euro gap between the cost of the two programs would be closed in 16 to 23 months. When considering gains for the "society as a whole," that is, the sum of the budget's savings and the value from the creation of jobs, one can see that the new SE program could make up for its higher cost in just 9 to 11 months.

There is of course no guarantee that the estimated effects will persist beyond the 18-month follow-up period. However, Figures 4 and 6 do not suggest that the control group is catching up with the treatment group and that the gap would close quickly. Thus, we argue that the estimated time that SE needs to make up for its higher cost (between 9 and 23 months depending on the chosen perspective) is relatively low compared to the potential future benefits.

6 Conclusion

The growing number of Disability Insurance (DI) recipients with mental health conditions, who often suffer ups and downs in the evolution of their illness, has blurred the line between those who are totally and permanently disabled and those who retain some work capacity or could recover it in the future. This trend motivates the implementation of active labor market programs to help DI beneficiaries return to work when their health allows it.

In this paper, we study the effects a new Supported Employment (SE) program introduced in March 2018 in Belgium. The program is characterized by a "work-first" approach with intense job coaching and follow-along support. Using a Randomized Control Trial, we compare the effects of this newly introduced program with regular Vocational Rehabilitation (VR) services, which have been in place for more than a decade in Belgium (i.e., control condition). Between March 2018 and December 2019, we recruited more than 660

DI recipients who suffer from mental illnesses and were willing to take part in a return-to-work program. Participants from across Belgium were randomly assigned to the new SE program or the regular VR program. We followed them for 18 months from the start of their program using both data from administrative registers and survey instruments designed for this research.

We find that compared to regular rehabilitation, SE increases the probability that DI recipients with mental conditions work while on claim and reduces their reliance on DI benefits. Specifically, we estimate that 18 months after the start of their return-to-work program, participants in the SE group are 9.5 percentage points more likely to be working part-time and receive 6% less in DI benefits than those in the control group. In addition, we observe that they find occupations in the private sector (and not in sheltered workshops).

We take advantage of our rich survey data to explore the channels through which SE achieves higher re-employment rates than regular rehabilitation. We find that participants in SE dedicate more time to the job search and less time to vocational training. In addition, DI recipients in SE report that they seek a job with fewer working hours and with lower pay. We hypothesize that SE participants form more realistic expectations of their capacity to work and as such are more likely to look for part-time jobs. In contrast, those in the control group who spend more time on vocational training might seek full work resumption, which is not always compatible with their current health status.

These findings should be of broad interest outside of Belgium since most OECD countries face rising disability rolls, especially individuals with mental health conditions. In this paper, we show that SE is successful in increasing the proportion of DI recipients who work while on claim, therefore reducing their reliance on benefits and easing the burden for the social security budget. The cost-benefit analysis reveals that the higher cost of SE per individual participant can be compensated within a reasonable time frame (of less than two years) if the observed effects were to remain beyond the 18-month follow-up window. This is of course an assumption that needs to be tested in future research.

References

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association*, *103*(484), 1481-1495.

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

Barrera-Osorio, F., Kugler, A. D., & Silliman, M. I. (2021). Job training through turmoil. *National Bureau of Economic Research Working Paper Series*, *No. 29565*.

Beuermann, D. W., Bottan, N. L., Hoffmann, B., Jackson, C. K., & Vera Cossio, D. A. (2021). Does education prevent job loss during downturns? evidence from exogenous school assignments and covid-19 in barbados. *National Bureau of Economic Research Working Paper Series*, *No. 29231*.

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

Bound, J., & Burkhauser, R. V. (1999). Economic analysis of transfer programs targeted on people with disabilities. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3, p. 3417-3528). New York: Elsevier Science.

Broadway, B., & Kassenboehmer, S. C. (2019). Employment effects of job counseling for disability insurance recipients. *Melbourne Institute Working Paper*, *19*(18).

Card, D., Kluve, J., & Weber, A. (2018). What works? a meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, *16*(3), 894-931.

Cohen, J., & Dupas, P. (2010). Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment*. *The Quarterly Journal of Economics*, *125*(1), 1-45.

Cottier, L., Flückiger, Y., Kempeneers, P., & Lalive, R. (2018). Does job search assistance really raise employment? *IZA Discussion paper*(11766).

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., & Zamora, P. (2013). Do labor

market policies have displacement effects? evidence from a clustered randomized experiment *. *The Quarterly Journal of Economics*, *128*(2), 531-580.

Crépon, B., Gurgand, M., Kamionka, T., & Lequien, L. (2012, Feb). *Is counseling welfare recipients cost-effective ? lessons from a random experiment.* Center for Research in Economics and Statistics.

Dean, D., Pepper, J., Schmidt, R., & Stern, S. (2017). The effects of vocational rehabilitation services for people with mental illness. *Journal of Human Resources*, *52*(3), 826-858.

Drake, R. E., McHugo, G. J., Becker, D. R., Anthony, W. A., & Clark, R. E. (1996). The new hampshire study of supported employment for people with severe mental illness. *Journal of Consulting and Clinical Psychology*, *64*(2), 391-399.

Eurofound. (2020). Living, working and covid-19. *Publications Office of the European Union*, *COVID-19 series*.

Field, E. M., Linden, L. L., Malamud, O., Rubenson, D., & Wang, S.-Y. (2019). Does vocational education work? evidence from a randomized experiment in mongolia. *National Bureau of Economic Research Working Paper Series*, *No. 26092*.

Fisher, R. A. (1935). *The design of experiments*. Macmillan.

Fogelgren, M., Ornstein, P., Rödin, M., & Skogman Thoursie, P. (2021). Is supported employment effective for young adults with disability pension? evidence from a swedish randomized evaluation. *Journal of Human Resources*, *(forthcoming)*.

Fontenay, S., & Tojerow, I. (2020). Work disability after motherhood and how paternity leave can help. *IZA Discussion Paper*(13756).

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

Frey, W. D., Drake, R. E., Bond, G. R., Miller, A. L., Goldman, H. H., Salkever, D. S., & Holsenbeck, S. (2011). *Mental health treatment study: Final report to social security administration* (Tech. Rep.).

Fujiwara, T., & Wantchekon, L. (2013). Can informed public deliberation overcome clientelism? experimental evidence from benin. *American Economic Journal: Applied*

Economics, *5*(4), 241-55.

Gerritse, E., Marcato, A., Plasman, R., & Tojerow, I. (2017). *Incapacité de travail: Quelles transitions et trajectoires après l'épisode d'incapacité?*

Hainmueller, J., & Xu, Y. (2013). ebalance: A stata package for entropy balancing. *Journal of Statistical Software*, *54*(7), 1 - 18.

Heß, S. (2017). Randomization inference with stata: A guide and software. *Stata Journal*, *17*(3), 630-651.

Jones, F., Mandy, A., & Partridge, C. (2009). Changing self-efficacy in individuals following a first time stroke: preliminary study of a novel self-management intervention. *Clinical Rehabilitation*, *23*(6), 522-533.

Kessler, R. C., Akiskal, H. S., Ames, M., Birnbaum, H., Greenberg, P., Hirschfeld, R. M., ... Wang, P. S. (2006). Prevalence and effects of mood disorders on work performance in a nationally representative sample of u.s. workers. *Am J Psychiatry*, *163*(9), 1561-8.

Kessler, R. C., Heeringa, S., Lakoma, M. D., Petukhova, M., Rupp, A. E., Schoenbaum, M., ... Zaslavsky, A. M. (2008). Individual and societal effects of mental disorders on earnings in the united states: results from the national comorbidity survey replication. *Am J Psychiatry*, *165*(6), 703-11.

Laun, L., & Thoursie, P. S. (2014). Does privatisation of vocational rehabilitation improve labour market opportunities? evidence from a field experiment in sweden. *Journal of Health Economics*, *34*, 59-72.

Liebman, J. B. (2015). Understanding the increase in disability insurance benefit receipt in the united states. *Journal of Economic Perspectives*, *29*(2), 123-50.

Luciano, A., Drake, R. E., Bond, G. R., Becker, D. R., Carpenter-Song, E., Lord, S., ... Swanson, S. J. (2014). Evidence-based supported employment for people with severe mental illness: Past, current, and future research. *Journal of Vocational Rehabilitation*, *40*(1), 1-13.

Maestas, N., Mullen, K. J., & Strand, A. (2013). Does disability insurance receipt discourage work? using examiner assignment to estimate causal effects of ssdi receipt.

American Economic Review, *103*(5), 1797-1829.

Markussen, S., & Røed, K. (2014). The impacts of vocational rehabilitation. *Labour Economics*, *31*, 1-13.

Marshall, T., Goldberg, R. W., Braude, L., Dougherty, R. H., Daniels, A. S., Ghose, S. S., ... Delphin-Rittmon, M. E. (2014). Supported employment: assessing the evidence. *Psychiatr Serv*, *65*(1), 16-23.

Modini, M., Tan, L., Brinchmann, B., Wang, M. J., Killackey, E., Glozier, N., ... Harvey, S. B. (2016). Supported employment for people with severe mental illness: systematic review and meta-analysis of the international evidence. *Br J Psychiatry*, *209*(1), 14-22.

OECD. (2010). *Sickness, disability and work: Breaking the barriers.*

Rosenbaum, P. R. (2002). *Observational studies.* New York: Springer.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Ryan, P. (1998). Random allocation of treatments in blocks. *Stata Technical Bulletin*, *41*, 43-46.

Schwarzer, R. (1992). *Self-efficacy: Thought control of action.* Washington, DC, US: Hemisphere Publishing Corp.

Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In . M. J. J. Weinman S. Wright (Ed.), *Measures in health psychology: A user's portfolio* (p. 35-37). Windsor, UK: NFER-NELSON.

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*, *34*(3), 220-233.

Wewiorski, N. J., & Fabian, E. S. (2004). Association between demographic and diagnostic factors and employment outcomes for people with psychiatric disabilities: A synthesis of recent research. *Mental Health Services Research*, *6*(1), 9-21.

Table 1: Characteristics of the Two Return-To-Work Programs

	Treatment	Control
Program:	Supported Employment (SE)	Vocational Rehabilitation (VR)
Launch:	2018	2009
Focus:	"Work first" approach with rapid job search	Human capital accumulation through training
Intensity:	1 caseworker for 20 beneficiaries (contact at least every 2 weeks)	1 caseworker for 100 beneficiaries
Duration:	Unlimited follow-along support	Max. 6 months after training completed

Table 2: Descriptive Statistics and Balancing Test

Panel A	Sample	Control Mean (SD)	Treatment	Balancing test (T-C) Coeff. (SE)
Administrative data				
Female (0/1)	0.50	0.51	0.49	-0.02
	(0.50)	(0.50)	(0.50)	(0.04)
Age	40.47	41.02	40.03	-1.00
	(8.50)	(8.43)	(8.54)	(0.66)
Live in Flanders (0/1)	0.69	0.70	0.69	-0.01
	(0.46)	(0.46)	(0.46)	(0.04)
Months on disability (#)	44.11	44.32	43.94	-0.38
	(37.34)	(36.16)	(38.31)	(2.89)
Long-term disability (0/1)	0.83	0.84	0.83	-0.02
	(0.37)	(0.37)	(0.38)	(0.03)
Dependents - children or spouse (0/1)	0.20	0.22	0.18	-0.04
	(0.40)	(0.42)	(0.38)	(0.03)
Blue collar worker (0/1)	0.53	0.55	0.51	-0.05
	(0.50)	(0.50)	(0.50)	(0.04)
Daily benefits (euros)	47.24	48.00	46.63	-1.38
	(11.84)	(12.22)	(11.51)	(0.93)
Voluntary work (0/1)	0.16	0.15	0.16	0.01
	(0.37)	(0.36)	(0.37)	(0.03)
Baseline survey				
Belgian nationality (0/1)	0.87	0.87	0.88	0.01
	(0.33)	(0.34)	(0.33)	(0.03)
Higher education (0/1)	0.26	0.24	0.27	0.03
	(0.44)	(0.43)	(0.45)	(0.03)
Number of observations	667	298	369	667
Panel B				
F-test regression of treatment on all outcomes	0.79			
P-value	0.65			

*Notes: Panel A reports descriptive statistics before program assignment for the sample of individuals who entered the study between March 2018 and December 2019. Columns "Sample", "Control" and "Treatment" report the means (standard deviations in parentheses) of individual characteristics for the whole sample, the treatment and the control sub-samples, respectively. Column "Balancing test" reports the difference (standard errors in parentheses) between those assigned to treatment and those assigned to control for each outcome considered. Panel B reports an aggregate test for the equality of means across all variables. Data sources are from NIHDI administrative registers, as well as baseline survey administered to all participants before their entry in a return-to-work program. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 3: Response to Follow-up Surveys and Attrition Test

	Attrition (0/1)		
	Follow-up 6	Follow-up 12	Follow-up 18
Panel A			
Treatment (0/1)	-0.041	-0.063	0.009
	(0.037)	(0.039)	(0.039)
N	667	667	667
Attrition mean	0.360	0.448	0.508
Panel B			
Treatment (0/1)	-0.039	-0.056	0.010
	(0.037)	(0.038)	(0.038)
Female (0/1)	-0.156 ***	-0.101 **	-0.112 ***
	(0.037)	(0.039)	(0.040)
Age	-0.004 *	-0.001	-0.006 **
	(0.002)	(0.002)	(0.002)
Live in Flanders (0/1)	0.085 **	-0.008	0.023
	(0.039)	(0.042)	(0.042)
Months on disability (#)	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)
Long-term disability (0/1)	-0.046	-0.022	-0.013
	(0.052)	(0.056)	(0.057)
Dependents - children or spouse (0/1)	-0.045	-0.061	-0.065
	(0.055)	(0.057)	(0.056)
Blue collar worker (0/1)	0.057	0.087 **	0.098 **
	(0.040)	(0.041)	(0.042)
Disability benefits (euros)	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)
Voluntary work (0/1)	-0.070	-0.076	-0.061
	(0.052)	(0.054)	(0.056)
Belgian nationality (0/1)	-0.120 **	-0.074	-0.119 **
	(0.059)	(0.058)	(0.059)
Higher education (0/1)	-0.174 ***	-0.187 ***	-0.111 **
	(0.043)	(0.046)	(0.048)
N	667	667	667
Panel C			
Treatment (0/1)	0.034	0.212	0.213
	(0.281)	(0.293)	(0.296)
Baseline characteristics + Interactions with treatment			
N	667	667	667
Aggregate F-test	0.83	1.16	0.39
F test: p-value joint significance of interactions	0.62	0.31	0.97

*Notes: This table presents an analysis of survey attrition for the follow-up at 6, 12 and 18 months. The dependent variable is a dummy that takes on the value 1 if the person did not answer the follow-up survey. Panel A presents the difference in response rate between the treatment and control groups. Panel B presents an analysis on the type of people that were less likely to be surveyed. The covariates include the treatment indicator, as well as all the variables reported in Table 2. Panel C presents a test of whether the treatment affected the type of person who completed the follow-up surveys, in other words whether the treatment caused a sample composition bias. Most importantly, it reports the p-values for the joint significance of a full set of baseline characteristics interacted with the treatment indicator. Data sources are NIHDI administrative registers, as well as baseline survey administered to all participants before their entry in a return-to-work program. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 4: Administrative Registers - Status and Benefits

	(1) Disability status (0/1)	(2) Work while on claim (0/1)	(3) Training while on claim (0/1)	(4) Monthly DI benefits (euros)
Treat. 6 months	-0.019	0.029	-0.018	-64.129 **
	(0.014)	(0.023)	(0.013)	(28.952)
Obs.	667	667	667	667
Cont. mean	0.973	0.094	0.037	1181
Treat. 12 months	-0.011	0.061 **	-0.040 ***	-67.735 **
	(0.021)	(0.027)	(0.014)	(33.192)
Obs.	667	667	667	667
Cont. mean	0.930	0.121	0.050	1147
Treat. 18 months	-0.026	0.095 ***	-0.025 **	-69.967 **
	(0.021)	(0.027)	(0.011)	(35.470)
Obs.	667	667	667	667
Cont. mean	0.933	0.111	0.030	1165
Baseline cont.	YES	YES	YES	YES

*Notes: The table reports intention-to-treat effects - coefficient β in equation (1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 5: Survey on Employment, Type of Contract and Sector of Activity

	(1) Paid work (0/1)	(2) Permanent contract (0/1)	(3) Temporary contract (0/1)	(4) Self-employed (0/1)	(5) Private sector (0/1)	(6) Public sector (0/1)	(7) Nonprofit sector (0/1)	(8) Sheltered work (0/1)
Treat. 6 months	0.055 **	0.035 *	0.031 **	-0.011	0.032	0.004	0.010	0.004
	(0.024)	(0.019)	(0.015)	(0.008)	(0.022)	(0.004)	(0.010)	(0.004)
Obs.	419	419	419	419	418	418	418	418
Cont. mean	0.044	0.022	0.011	0.011	0.039	0.000	0.006	0.000
Treat. 12 months	0.070 *	0.064 **	0.008	-0.002	0.054 *	-0.000	0.014 *	0.002
	(0.036)	(0.026)	(0.026)	(0.008)	(0.031)	(0.014)	(0.008)	(0.013)
Obs.	364	364	364	364	364	364	364	364
Cont. mean	0.104	0.039	0.058	0.006	0.071	0.019	0.000	0.013
Treat. 18 months	0.087 **	0.026	0.058 **	0.003	0.100 ***	0.001	0.009	-0.023
	(0.042)	(0.034)	(0.025)	(0.014)	(0.036)	(0.016)	(0.012)	(0.015)
Obs.	321	321	321	321	321	321	321	321
Cont. mean	0.130	0.089	0.027	0.014	0.075	0.021	0.007	0.027
Baseline cont.	NO	NO	NO	NO	NO	NO	NO	NO

*Notes: The table reports intention-to-treat effects - coefficient β in equation (1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from a survey specifically designed for this study. Participants were asked to fill out the questionnaire sent every 6 months by email or post (depending on their preference). Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 6: Survey on Earnings

	(1) Total earnings (euros)	(2) Wage income (euros)	(3) DI benefits (euros)	(4) UI benefits (euros)
Treat. 6 months	17.670	62.411	-56.923	12.182
	(49.220)	(39.443)	(44.466)	(12.788)
Obs.	405	405	405	405
Cont. mean	1203.379	78.784	1117.124	7.471
Treat. 12 months	17.018	44.226	-30.372	3.164
	(52.683)	(53.599)	(52.539)	(17.010)
Obs.	349	349	349	349
Cont. mean	1243.051	164.934	1051.763	26.354
Treat. 18 months	29.336	119.394 **	-104.797 **	14.739
	(56.851)	(54.140)	(52.833)	(15.415)
Obs.	316	316	316	316
Cont. mean	1239.128	132.379	1093.132	13.617
Baseline cont.	NO	NO	NO	NO

*Notes: The table reports intention-to-treat effects - coefficient β in equation (1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from a survey specifically designed for this study. Participants were asked to fill out the questionnaire sent every 6 months by email or post (depending on their preference). Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 7: Survey on Health and Well-being

	(1) Health related quality of life index	(2) Self-esteem index	(3) Self-Efficacy index
Treat. 6 months	0.016	0.036	0.021
	(0.052)	(0.080)	(0.082)
Obs.	407	394	394
Cont. mean	-0.071	-0.032	-0.020
Treat. 12 months	-0.030	-0.094	-0.028
	(0.061)	(0.091)	(0.091)
Obs.	352	345	345
Cont. mean	-0.014	0.116	0.045
Treat. 18 months	0.037	0.018	0.057
	(0.065)	(0.092)	(0.100)
Obs.	313	306	306
Cont. mean	-0.066	-0.008	0.002
Baseline cont.	YES	YES	YES

*Notes: The table reports intention-to-treat effects - coefficient β in equation (1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from a survey specifically designed for this study. Participants were asked to fill out the questionnaire sent every 6 months by email or post (depending on their preference). Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 8: Survey on Job Search Behavior and Vocational Training

	(1) Nb. meetings caseworker	(2) Search job (0/1)	(3) Time job search (hours)	(4) Preferred hours	(5) Preferred wage (euros)	(6) Training nb. of weeks
Treat. 6 months	4.804 ***	0.238 ***	0.545	-2.391 **	-123.851 **	0.041
	(0.471)	(0.047)	(1.014)	(1.031)	(56.706)	(0.566)
Obs.	405	411	395	388	382	409
Cont. mean	3.783	0.350	4.618	24.008	1399.697	1.958
Treat. 12 months	3.938 ***	0.237 ***	2.525 **	-2.969 ***	-143.158 **	-1.887 **
	(0.510)	(0.050)	(0.999)	(1.072)	(55.351)	(0.781)
Obs.	350	354	349	344	329	351
Cont. mean	2.293	0.248	2.666	25.108	1441.353	3.748
Treat. 18 months	3.063 ***	0.159 ***	0.700	-1.672	-170.560 **	-1.948 **
	(0.616)	(0.053)	(0.723)	(1.198)	(67.322)	(0.875)
Obs.	319	320	317	307	301	318
Cont. mean	1.896	0.266	2.593	23.485	1454.008	4.049
Baseline cont.	NO	NO	NO	NO	NO	NO

*Notes: The table reports intention-to-treat effects - coefficient β in equation (1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from a survey specifically designed for this study. Participants were asked to fill out the questionnaire sent every 6 months by email or post (depending on their preference). Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 9: The General Self-Efficacy Scale

	Not at all true (=1)	Hardly true (=2)	Moderately true (=3)	Exactly true (=4)
1) I can always manage to solve difficult problems if I try hard enough.	o	o	o	o
2) If someone opposes me, I can find the means and ways to get what I want.	o	o	o	o
3) It is easy for me to stick to my aims and accomplish my goals.	o	o	o	o
4) I am confident that I could deal efficiently with unexpected events.	o	o	o	o
5) Thanks to my resourcefulness, I know how to handle unforeseen situations.	o	o	o	o
6) I can solve most problems if I invest the necessary effort.	o	o	o	o
7) I can remain calm when facing difficulties because I can rely on my coping abilities.	o	o	o	o
8) When I am confronted with a problem, I can usually find several solutions.	o	o	o	o
9) If I am in trouble, I can usually think of a solution.	o	o	o	o
10) I can usually handle whatever comes my way.	o	o	o	o

Notes: English version of the Generalized Self-Efficacy Scale based on Schwarzer and Jerusalem (1995). The total score is calculated by finding the sum of the all items. For the composite score therefore ranges between 10 and 40, with a higher score indicating more self-efficacy.

Table 10: Impact of COVID-19 Pandemic on Probability to Work While on Claim (0/1)

	(1) Treatment	(2) Treatment x COVID19	(3) Difference
Follow-up 6	0.034	-0.031	0.004
	(0.024)	(0.041)	(0.038)
Follow-up 7	0.056 **	-0.047	0.009
	(0.025)	(0.039)	(0.035)
Follow-up 8	0.072 ***	-0.063 *	0.010
	(0.027)	(0.037)	(0.031)
Follow-up 9	0.101 ***	-0.096 ***	0.005
	(0.030)	(0.036)	(0.028)
Follow-up 10	0.130 ***	-0.116 ***	0.015
	(0.032)	(0.038)	(0.028)
Follow-up 11	0.147 ***	-0.134 ***	0.013
	(0.034)	(0.038)	(0.026)
Follow-up 12	0.155 ***	-0.115 ***	0.040
	(0.037)	(0.042)	(0.027)
Follow-up 13	0.180 ***	-0.150 ***	0.030
	(0.041)	(0.044)	(0.025)
Follow-up 14	0.166 ***	-0.111 **	0.055 **
	(0.044)	(0.047)	(0.026)
Follow-up 15	0.152 ***	-0.082	0.070 ***
	(0.047)	(0.050)	(0.026)
Follow-up 16	0.142 ***	-0.070	0.072 ***
	(0.053)	(0.056)	(0.025)
Follow-up 17	0.121 **	-0.036	0.085 ***
	(0.058)	(0.061)	(0.025)
Follow-up 18	0.140 **	-0.034	0.106 ***
	(0.069)	(0.071)	(0.026)
Obs.	12,673		

*Notes: The table reports intention-to-treat effects at different follow-up period from 6 to 18 months since the start of the return-to-work program. Column (1) displays pre-pandemic effects of SE (β in equation (3)). Column (2) displays the interaction between the treatment indicator and a binary indicator "COVID-19" that takes on a value 1 from the start of the pandemic in March 2020 (δ in equation (3)) and should be interpreted as the effect of the pandemic on the treated. Column (3) reports the difference between the first two columns and should be interpreted as the effect of SE for participants impacted by the COVID-19 pandemic. Standard errors in parentheses are clustered at the individual level. The OLS regression controls for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years). Compared to previous estimations, the sample now pulls together all participants who entered the study between March 2018 and December 2019 at each follow-up period (i.e. a panel of 667 participants over 18 time periods). Data source is from NIHDI administrative registers. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 11: Robustness Checks

	(1) Disability status (0/1)	(2) Work while on claim (0/1)	(3) Training while on claim (0/1)	(4) Monthly DI benefits (euros)	(5) Paid work (0/1)	(6) Permanent contract (0/1)	(7) Temporary contract (0/1)	(8) Self-employed (0/1)	(9) Private sector (0/1)	(10) Public sector (0/1)	(11) Nonprofit sector (0/1)	(12) Sheltered work (0/1)
Treat. 6 months	-0.019	0.029	-0.018	-64.129 **	0.055 **	0.035 *	0.031 **	-0.011	0.032	0.004	0.010	0.004
SE	(0.014)	(0.023)	(0.013)	(28.952)	(0.024)	(0.019)	(0.015)	(0.008)	(0.022)	(0.004)	(0.010)	(0.004)
OLS p-value	0.194	0.219	0.170	0.027	0.026	0.064	0.045	0.152	0.144	0.320	0.291	0.320
Rand. inf. p-value	0.189	0.226	0.140	0.023	0.032	0.079	0.056	0.124	0.151	0.981	0.388	0.877
Sharpened q-value	0.324	0.331	0.318	0.128	0.128	0.181	0.141	0.309	0.309	0.434	0.434	0.434
Obs.	667	667	667	667	419	419	419	419	418	418	418	418
Treat. 12 months	-0.011	0.061 **	-0.040 ***	-67.735 **	0.070 *	0.064 **	0.008	-0.002	0.054 *	-0.000	0.014 *	0.002
SE	(0.021)	(0.027)	(0.014)	(33.192)	(0.036)	(0.026)	(0.026)	(0.008)	(0.031)	(0.014)	(0.008)	(0.013)
OLS p-value	0.582	0.023	0.004	0.042	0.050	0.015	0.753	0.806	0.083	0.975	0.085	0.863
Rand. inf. p-value	0.571	0.026	0.001	0.048	0.047	0.020	0.763	0.513	0.083	0.963	0.233	0.859
Sharpened q-value	0.698	0.053	0.036	0.072	0.079	0.045	0.760	0.760	0.117	0.880	0.117	0.760
Obs.	667	667	667	667	364	364	364	364	364	364	364	364
Treat. 18 months	-0.026	0.095 ***	-0.025 **	-69.967 **	0.087 **	0.026	0.058 **	0.003	0.100 ***	0.001	0.009	-0.023
SE	(0.021)	(0.027)	(0.011)	(35.470)	(0.042)	(0.034)	(0.025)	(0.014)	(0.036)	(0.016)	(0.012)	(0.015)
OLS p-value	0.216	0.001	0.023	0.049	0.040	0.454	0.020	0.829	0.006	0.935	0.468	0.141
Rand. inf. p-value	0.247	0.002	0.013	0.070	0.049	0.487	0.028	0.875	0.007	0.882	0.530	0.117
Sharpened q-value	0.231	0.007	0.067	0.085	0.085	0.415	0.067	0.564	0.036	0.598	0.415	0.165
Obs.	667	667	667	667	321	321	321	321	321	321	321	321
Baseline cont.	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	NO	NO

*Notes: The table reports intention-to-treat effects - coefficient β in equation (1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. Under the name "Rand. inf. p-value," we report p-values from the randomization inference procedure developed by Heß for Stata (Heß, 2017) based on 1,000 random draws. The test statistic reports the proportion of placebo coefficients that are larger than the actual treatment effect. Under the name "Sharpened q-value," we report p-values adjusted for multiple inference testing using the procedure by Anderson (2008). The latter controls for the false discovery rate, that is the expected proportion of rejections that are type I errors. Significance levels are based on classical p-values reported under the name "OLS p-value": *** p<0.01, ** p<0.05, * p<0.1*

Table 12: Robustness Checks (cont.)

	(13) Total earnings (euros)	(14) Wage income (euros)	(15) DI benefits (euros)	(16) UI benefits (euros)	(17) Health related quality of life index	(18) Self-esteem index	(19) Self-Efficacy index	(20) Nb. meetings case-worker	(21) Search job (0/1)	(22) Time job search (hours)	(23) Preferred hours	(24) Preferred wage (euros)	(25) Training nb. of weeks
Treat. 6 months	17.670	62.411	-56.923	12.182	0.016	0.036	0.021	4.804 ***	0.238 ***	0.545	-2.391 **	-123.851 **	0.041
SE	(49.220)	(39.443)	(44.466)	(12.788)	(0.052)	(0.080)	(0.082)	(0.471)	(0.047)	(1.014)	(1.031)	(56.706)	(0.566)
OLS p-value	0.720	0.114	0.201	0.341	0.765	0.649	0.798	0.000	0.000	0.591	0.021	0.030	0.942
Rand. inf. p-value	0.709	0.106	0.189	0.400	0.763	0.657	0.803	0.000	0.000	0.575	0.022	0.015	0.942
Sharpened q-value	0.563	0.276	0.324	0.440	0.576	0.561	0.576	0.001	0.001	0.561	0.128	0.128	0.605
Obs.	405	405	405	405	407	394	394	405	411	395	388	382	409
Treat. 12 months	17.018	44.226	-30.372	3.164	-0.030	-0.094	-0.028	3.938 ***	0.237 ***	2.525 **	-2.969 ***	-143.158 **	-1.887 **
SE	(52.683)	(53.599)	(52.539)	(17.010)	(0.061)	(0.091)	(0.091)	(0.510)	(0.050)	(0.999)	(1.072)	(55.351)	(0.781)
OLS p-value	0.747	0.410	0.564	0.853	0.625	0.303	0.762	0.000	0.000	0.012	0.006	0.010	0.016
Rand. inf. p-value	0.719	0.391	0.545	0.871	0.617	0.297	0.763	0.000	0.000	0.012	0.005	0.008	0.007
Sharpened q-value	0.760	0.488	0.698	0.760	0.716	0.352	0.760	0.001	0.001	0.044	0.036	0.044	0.045
Obs.	349	349	349	349	352	345	345	350	354	349	344	329	351
Treat. 18 months	29.336	119.394 **	-104.797 **	14.739	0.037	0.018	0.057	3.063 ***	0.159 ***	0.700	-1.672	-170.560 **	-1.948 **
SE	(56.851)	(54.140)	(52.833)	(15.415)	(0.065)	(0.092)	(0.100)	(0.616)	(0.053)	(0.723)	(1.198)	(67.322)	(0.875)
OLS p-value	0.606	0.028	0.048	0.340	0.571	0.843	0.571	0.000	0.003	0.334	0.164	0.012	0.027
Rand. inf. p-value	0.591	0.036	0.053	0.393	0.583	0.857	0.581	0.000	0.002	0.312	0.159	0.011	0.014
Sharpened q-value	0.435	0.067	0.085	0.351	0.428	0.564	0.428	0.001	0.024	0.351	0.180	0.053	0.067
Obs.	316	316	316	316	313	306	306	319	320	317	307	301	318
Baseline cont.	NO	NO	NO	NO	YES	YES	YES	NO	NO	NO	NO	NO	NO

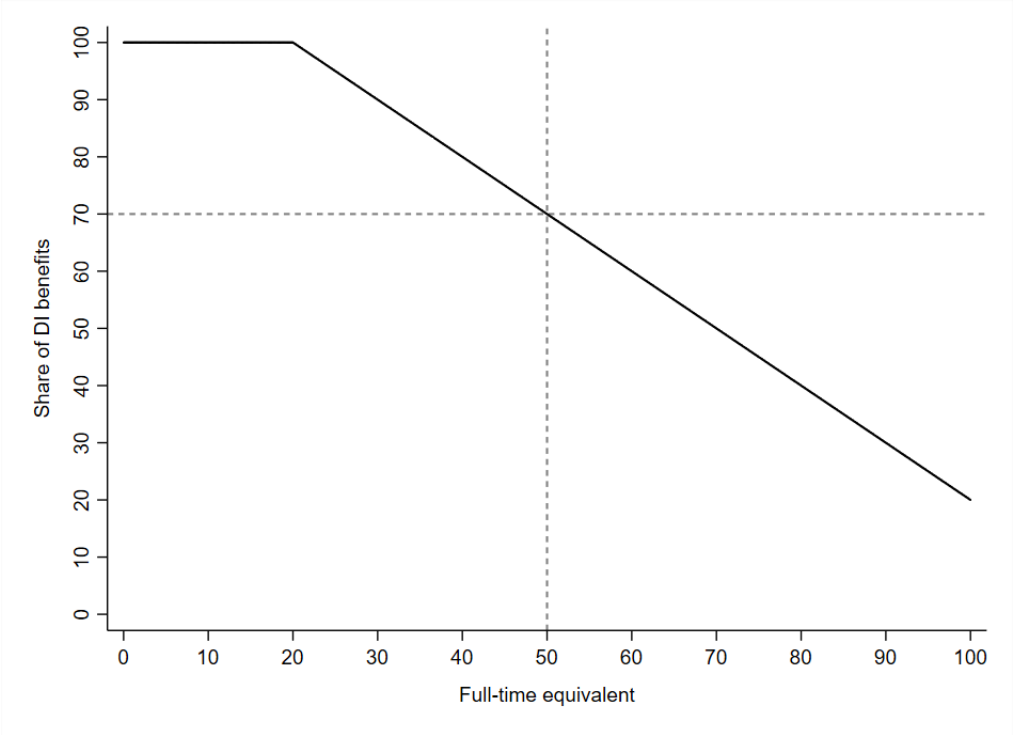
*Notes: The table reports intention-to-treat effects - coefficient β in equation (1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. Under the name "Rand. inf. p-value," we report p-values from the randomization inference procedure developed by Heß for Stata (Heß, 2017) based on 1,000 random draws. The test statistic reports the proportion of placebo coefficients that are larger than the actual treatment effect. Under the name "Sharpened q-value," we report p-values adjusted for multiple inference testing using the procedure by Anderson (2008). The latter controls for the false discovery rate, that is the expected proportion of rejections that are type I errors. Significance levels are based on classical p-values reported under the name "OLS p-value": *** p<0.01, ** p<0.05, * p<0.1*

Table 12: Cost-Benefit Analysis

Cost per participant (18 months)	SE = 7200 euros	VR = 4800 euros
Survey sample	**Effect size / SE (euros)**	**Catch up time (months)**
Reduction in DI benefits paid by Social Security	-104.80 (52.83)	23
Increase in wage income for DI recipients	119.39 (54.14)	20
Gains for "Society as a whole"	224.19 (90.96)	11
Reweighted sample **(female=61%, flanders=50%, blue collar=59%)**	**Effect size / SE (euros)**	**Catch up time (months)**
Reduction in DI benefits paid by Social Security	-147.39 (58.55)	16
Increase in wage income for DI recipients	116.61 (56.76)	21
Gains for "Society as a whole"	264.00 (97.63)	9

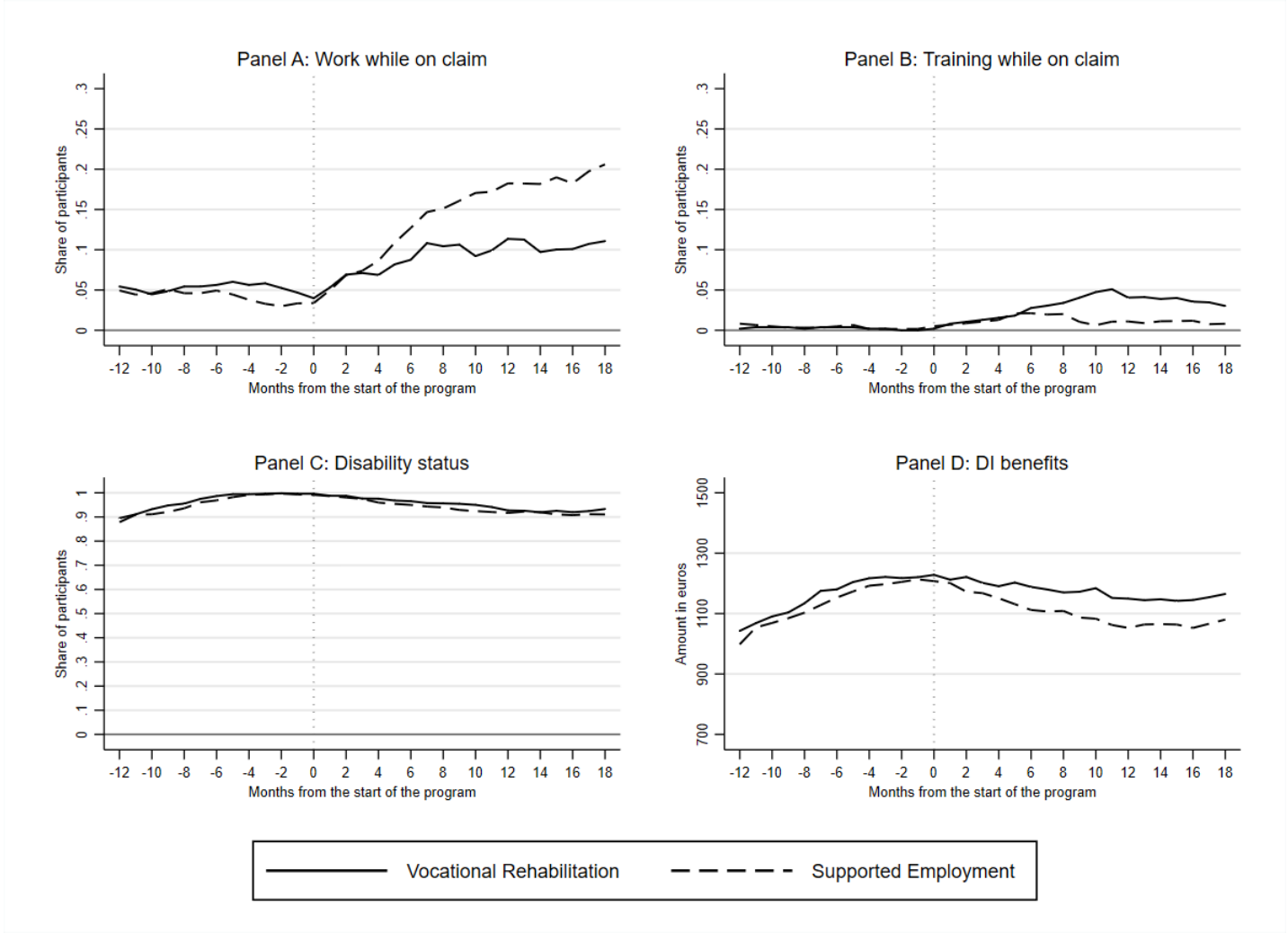
Notes: The cost per participant is laid out in a cooperation agreement between NIHDI and regional employment agencies who offer rehabilitation services. The table also reports intention-to-treat effects - coefficient β in equation (1) - from separate OLS regressions using survey answers at follow-up 18 months. The results for the "survey sample" are the same as in Table 6. We also use the entropy balancing method by Hainmueller & Xu (2013) to reweight the survey sample to known characteristics from the population of DI recipients with mental conditions in Belgium. The results using this balanced sample are reported under the title "reweighted sample." The outcome "gains for society as a whole" encompasses both the budget savings for NIHDI in the form of reduced benefits and the value of the production generated by the new jobs (estimated by the wage cost). It is therefore the sum of the absolute value of the two other outcomes. The "catch up time" corresponds to the number of months that the benefits of the SE program would need to last, beyond the 18-month follow-up period, to make up for its higher cost.

Figure 1: Reduction in Disability Insurance Benefits when Working while on Claim



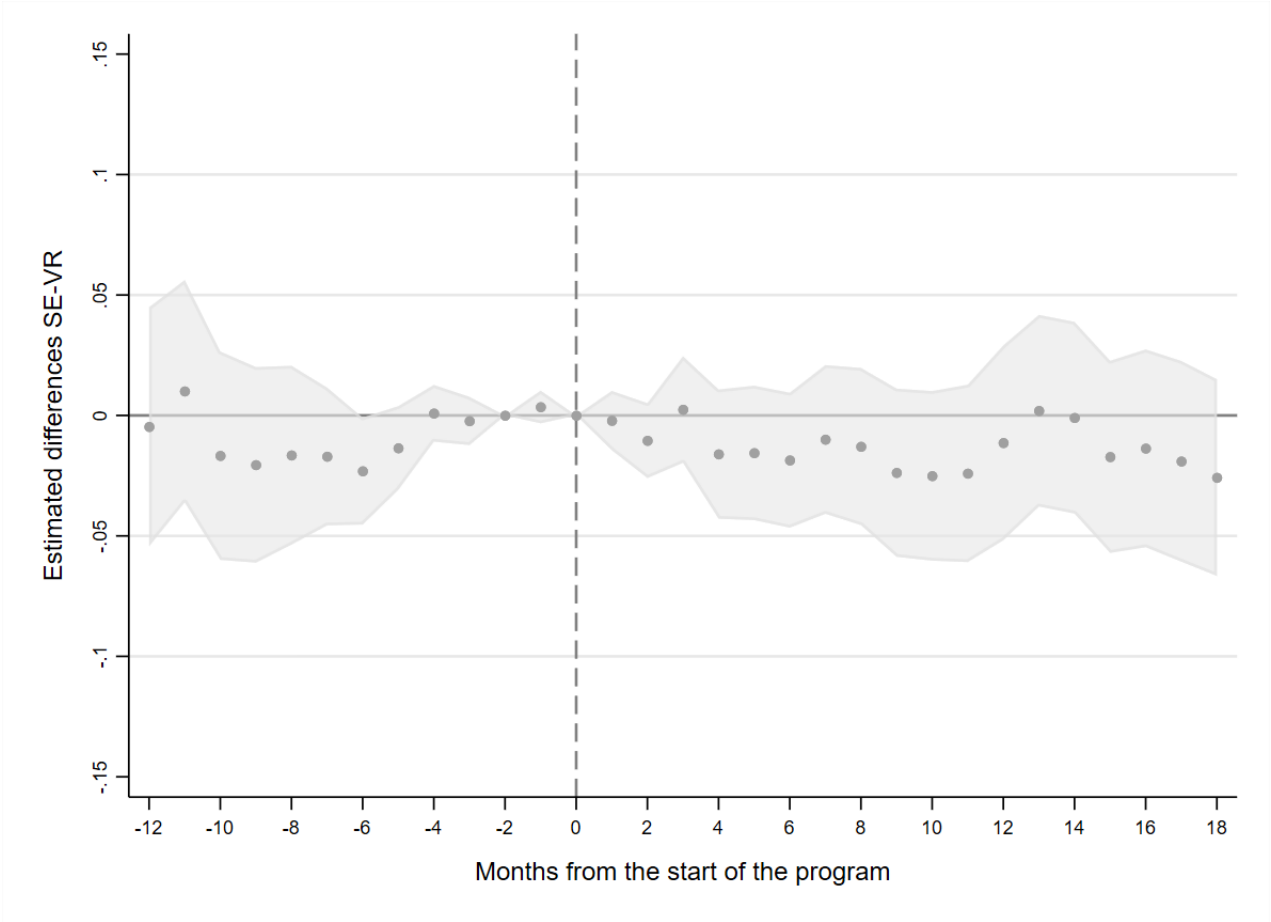
Notes: The rules are set in the Royal Decree implementing the law on compulsory insurance for medical care and cash benefits, consolidated on 14 July 1994, and amended in February 2018. Disability Insurance benefits are reduced by the amount of working time that exceeds 20% of a Full Time Equivalent (that is 38 hours a week in Belgium). DI recipients who work 20% (or less) of a FTE keep their full benefits. DI recipients who work half-time (50% of FTE) keep 70% of their benefits.

Figure 2: Comparison of Participants by Months since Randomly Allocated to a Return-to-work Program



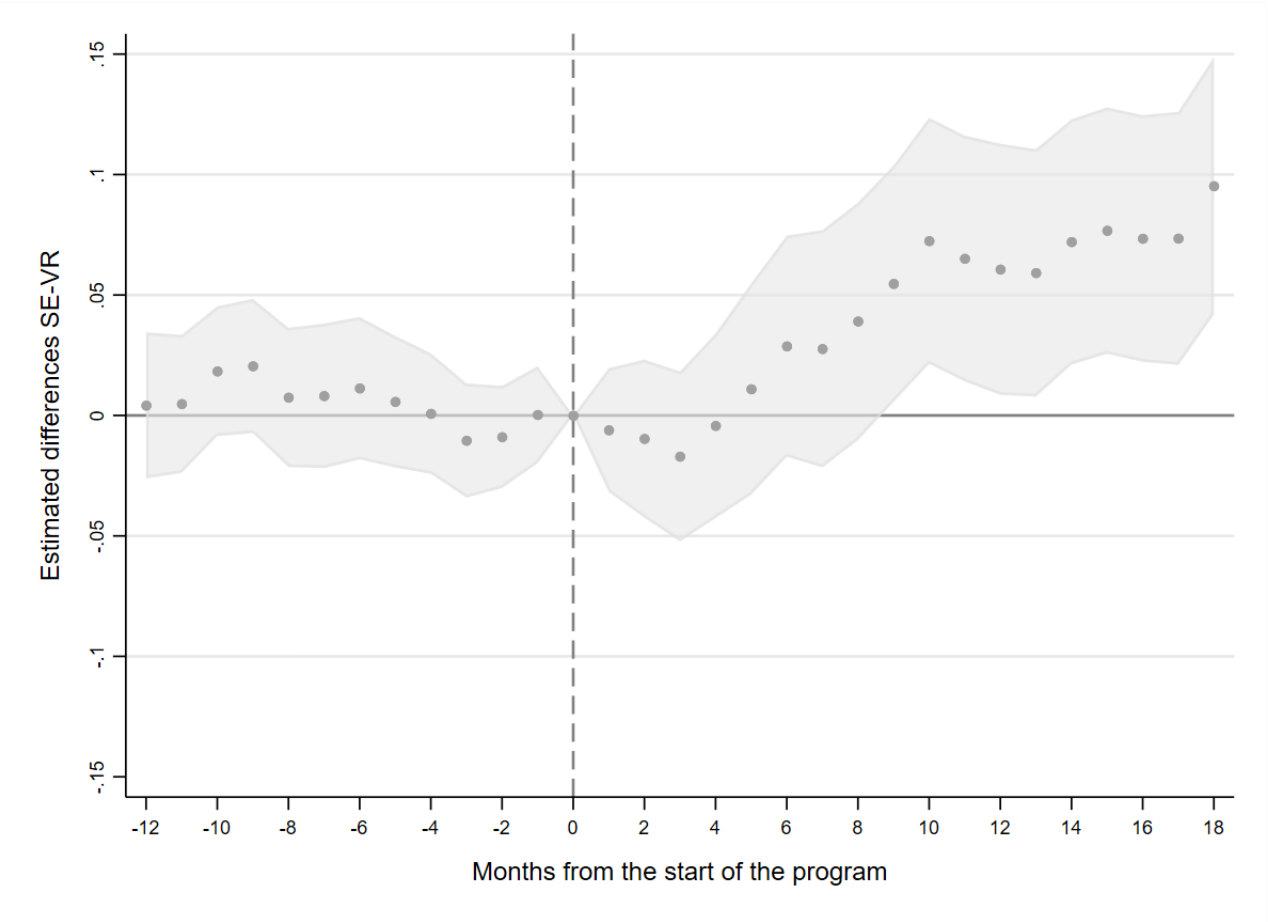
Notes: Horizontal axes show months since filling out baseline questionnaire and starting the return-to-work program. Vertical axes plot unconditional means for the share of study participants who work while on claim (Panel A), follow a training while on claim (Panel B), retain their disability status (Panel C), as well as the amount of monthly benefits they receive (Panel D). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.

Figure 3: Intention-to-treat Effects on Disability Status (0/1)



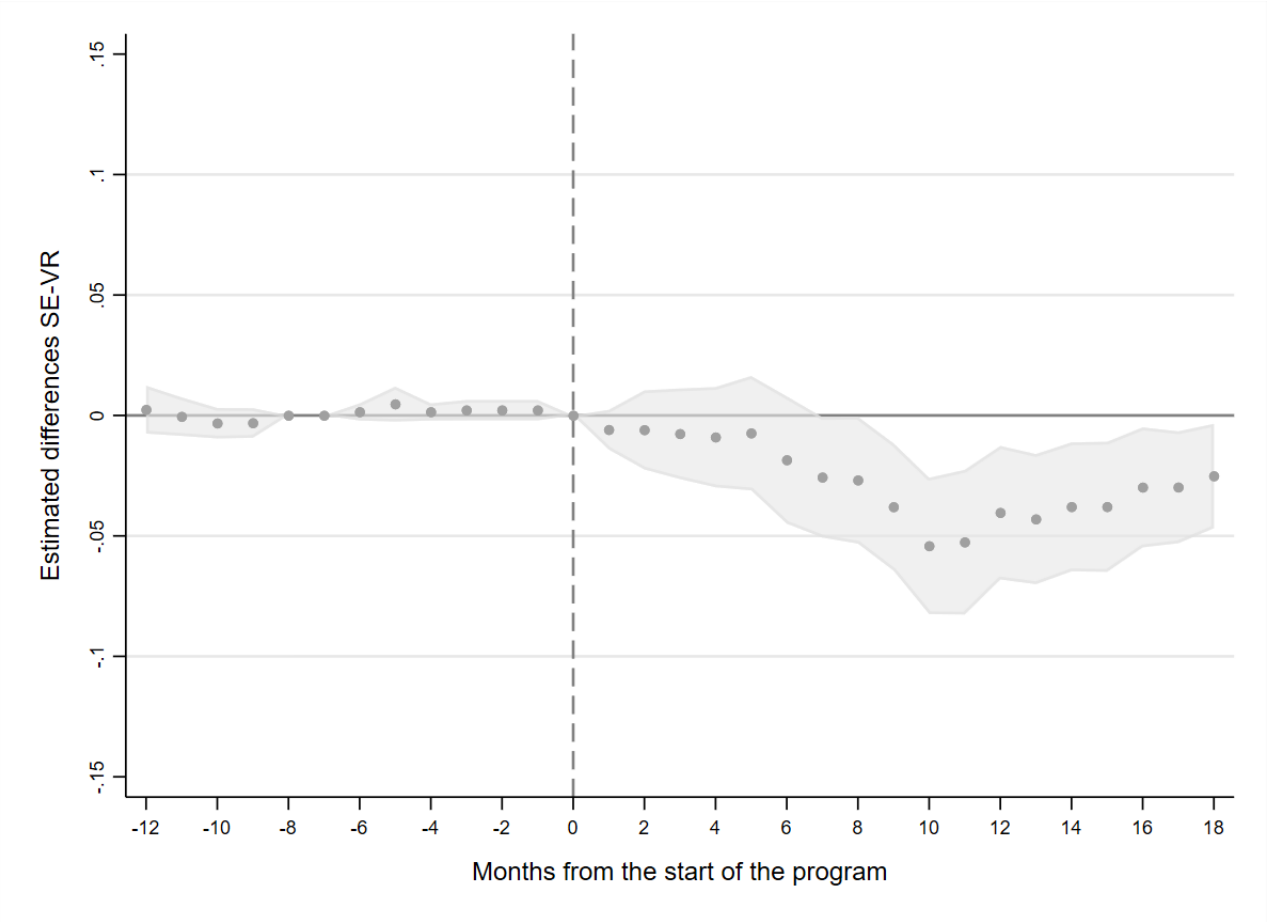
Notes: Horizontal axis shows months since filling out baseline questionnaire and starting the return-to-work program. Vertical axis plots the difference between Supported Employment (SE) and Vocational Rehabilitation (VR). Each dot denotes the point estimate for intention-to-treat effect - coefficient β in equation (1) - at a given time horizon based on separate OLS regression. The shaded area denotes 95% confidence interval for the corresponding point estimate from heteroskedastic-robust standard errors. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.

Figure 4: Intention-to-treat Effects on Work while on Claim (0/1)



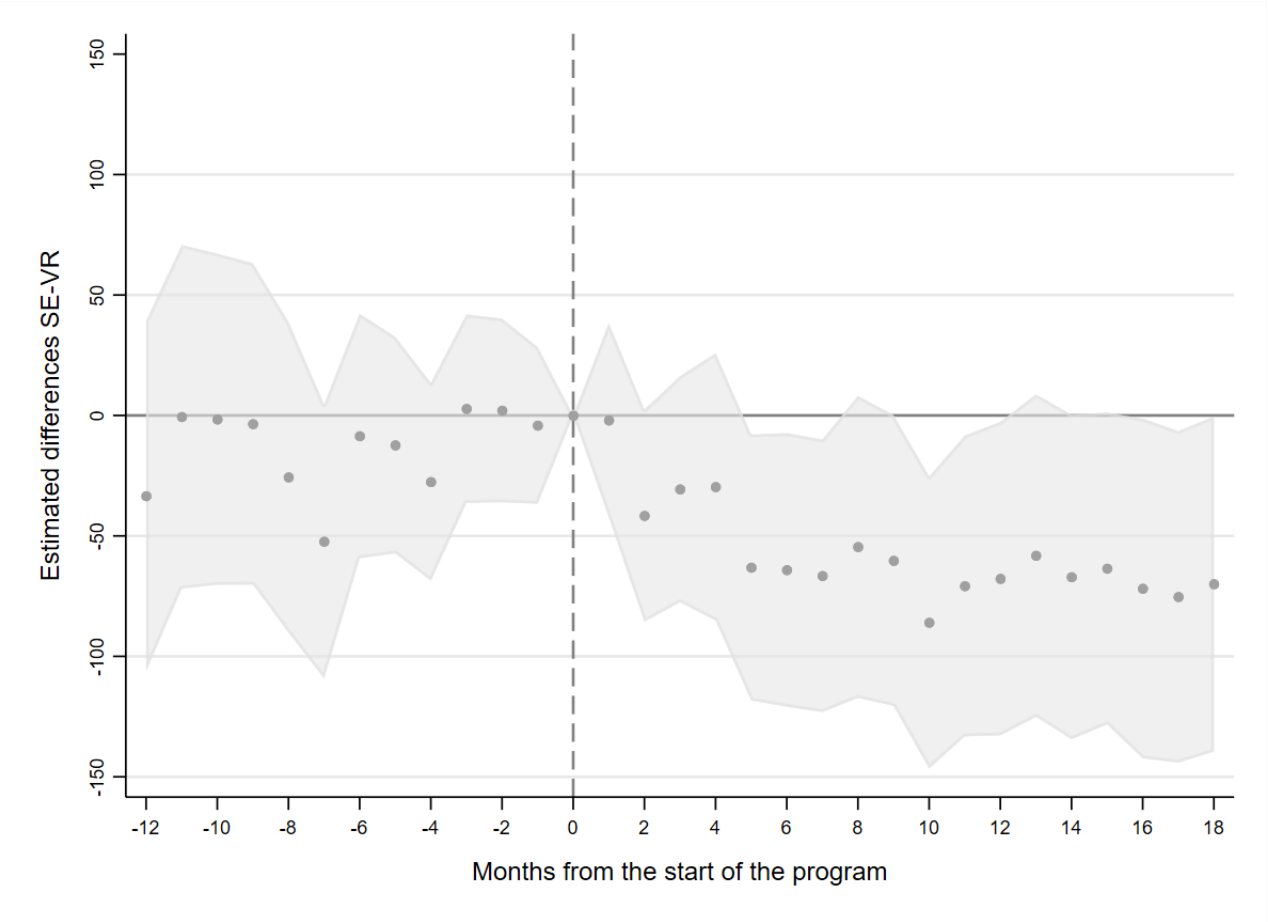
Notes: Horizontal axis shows months since filling out baseline questionnaire and starting the return-to-work program. Vertical axis plots the difference between Supported Employment (SE) and Vocational Rehabilitation (VR). Each dot denotes the point estimate for intention-to-treat effect - coefficient β in equation (1) - at a given time horizon based on separate OLS regression. The shaded area denotes 95% confidence interval for the corresponding point estimate from heteroskedastic-robust standard errors. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.

Figure 5: Intention-to-treat Effects on Training while on Claim (0/1)



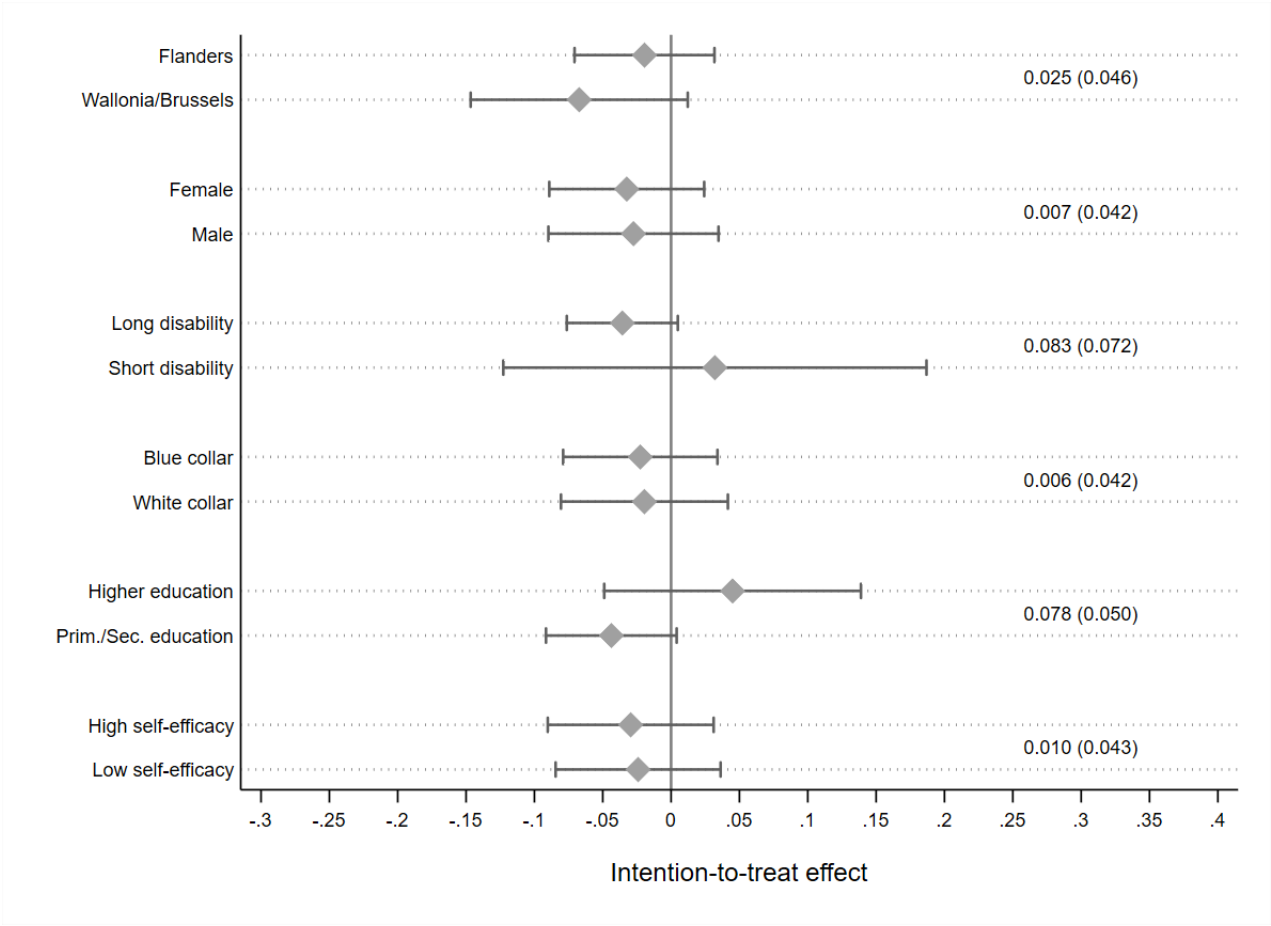
Notes: Horizontal axis shows months since filling out baseline questionnaire and starting the return-to-work program. Vertical axis plots the difference between Supported Employment (SE) and Vocational Rehabilitation (VR). Each dot denotes the point estimate for intention-to-treat effect - coefficient β in equation (1) - at a given time horizon based on separate OLS regression. The shaded area denotes 95% confidence interval for the corresponding point estimate from heteroskedastic-robust standard errors. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.

Figure 6: Intention-to-treat Effects on Disability Benefits (euros)



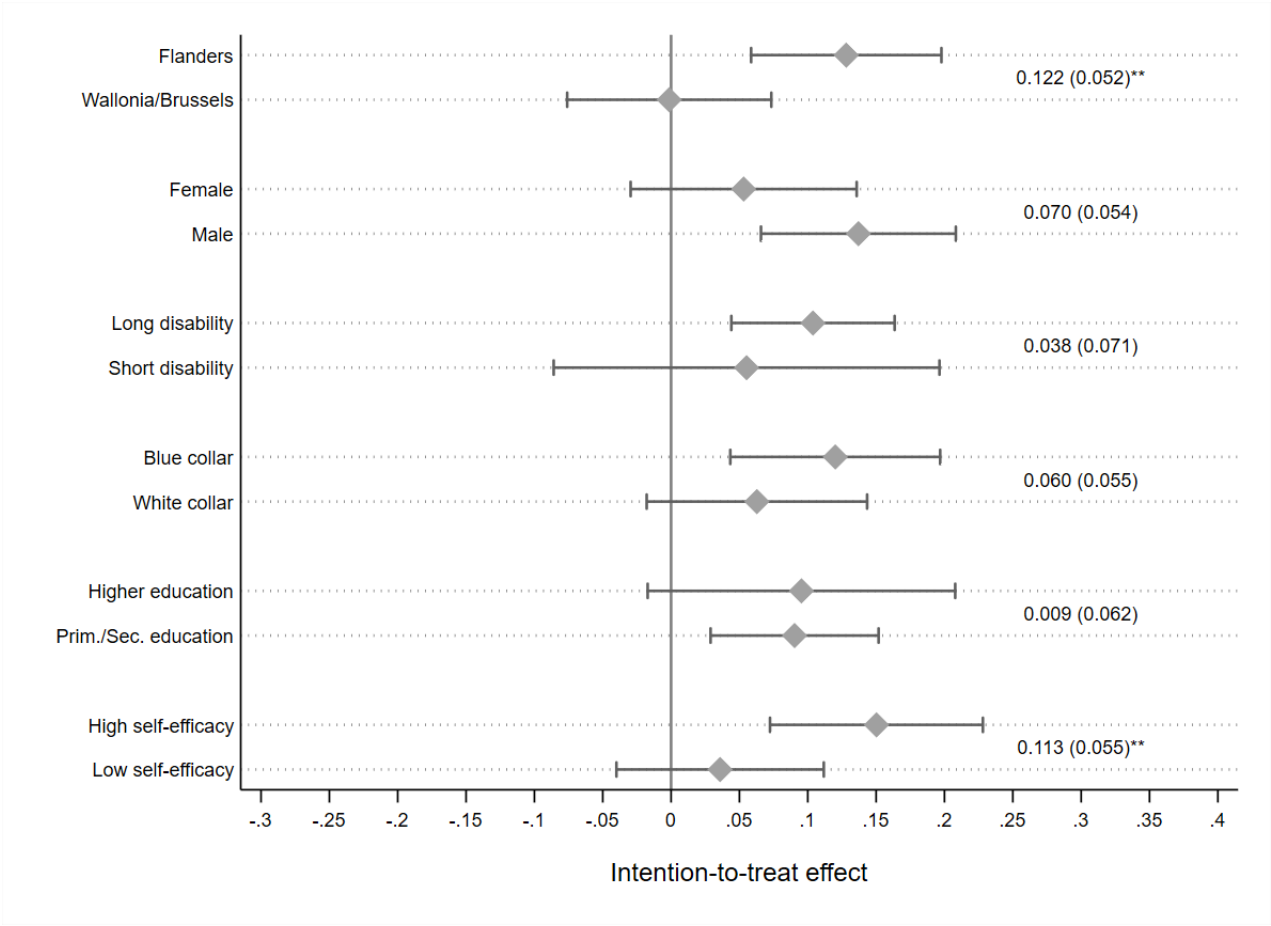
Notes: Horizontal axis shows months since filling out baseline questionnaire and starting the return-to-work program. Vertical axis plots the difference between Supported Employment (SE) and Vocational Rehabilitation (VR). Each dot denotes the point estimate for intention-to-treat effect - coefficient β in equation (1) - at a given time horizon based on separate OLS regression. The shaded area denotes 95% confidence interval for the corresponding point estimate from heteroskedastic-robust standard errors. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.

Figure 7: Heterogeneity Analysis: Effects on Disability Status (0/1)



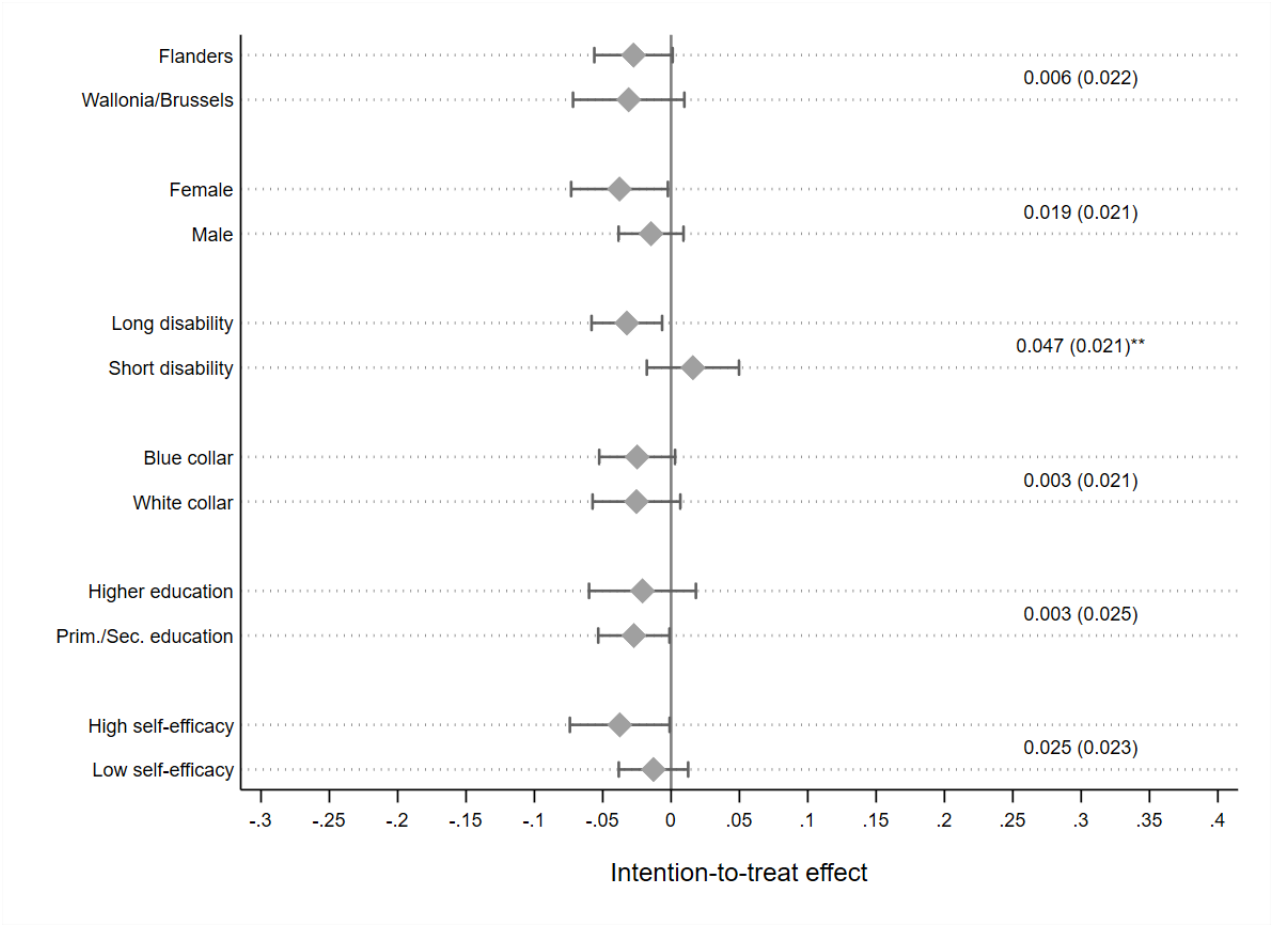
*Notes: The horizontal axis shows intention-to-treat estimates based on separate OLS regressions for the effects of Supported Employment (SE) compared to Vocational Rehabilitation (VR) 18 months after the start of the return-to-work program. The 95% confidence intervals are computed using heteroskedastic-robust standard errors. We also report on the graph the absolute value of the difference between the effects measured in the two groups (e.g. female vs male DI recipients), as well as the standard errors in parentheses (with significance levels displayed as *** p<0.01, ** p<0.05, * p<0.1). We use the same controls as in equation (1). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.*

Figure 8: Heterogeneity Analysis: Effects on Work while on Claim (0/1)



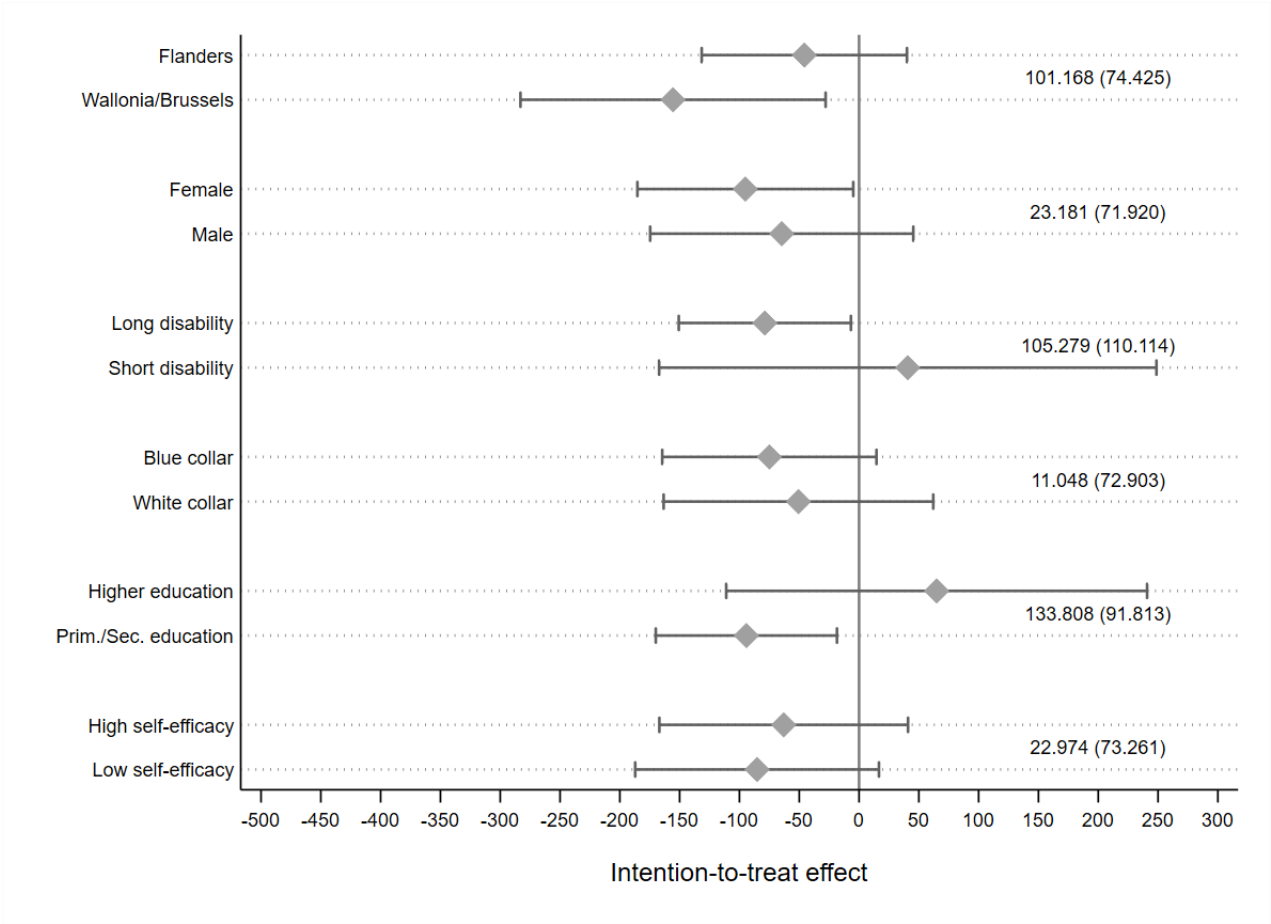
*Notes: The horizontal axis shows intention-to-treat estimates based on separate OLS regressions for the effects of Supported Employment (SE) compared to Vocational Rehabilitation (VR) 18 months after the start of the return-to-work program. The 95% confidence intervals are computed using heteroskedastic-robust standard errors. We also report on the graph the absolute value of the difference between the effects measured in the two groups (e.g. female vs male DI recipients), as well as the standard errors in parentheses (with significance levels displayed as *** p<0.01, ** p<0.05, * p<0.1). We use the same controls as in equation (1). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.*

Figure 9: Heterogeneity Analysis: Effects on Training while on Claim (0/1)



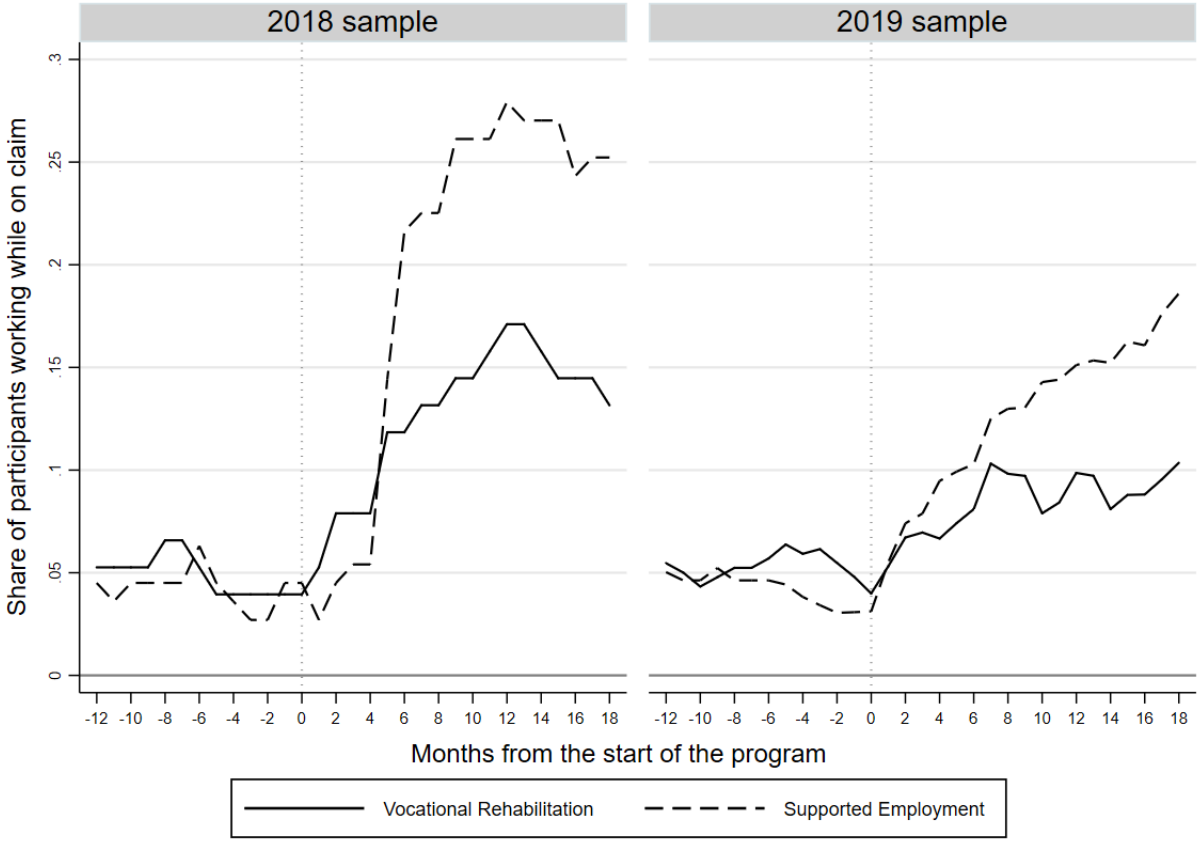
*Notes: The horizontal axis shows intention-to-treat estimates based on separate OLS regressions for the effects of Supported Employment (SE) compared to Vocational Rehabilitation (VR) 18 months after the start of the return-to-work program. The 95% confidence intervals are computed using heteroskedastic-robust standard errors. We also report on the graph the absolute value of the difference between the effects measured in the two groups (e.g. female vs male DI recipients), as well as the standard errors in parentheses (with significance levels displayed as *** p<0.01, ** p<0.05, * p<0.1). We use the same controls as in equation (1). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.*

Figure 10: Heterogeneity Analysis: Effects on Disability Benefits (euros)



*Notes: The horizontal axis shows intention-to-treat estimates based on separate OLS regressions for the effects of Supported Employment (SE) compared to Vocational Rehabilitation (VR) 18 months after the start of the return-to-work program. The 95% confidence intervals are computed using heteroskedastic-robust standard errors. We also report on the graph the absolute value of the difference between the effects measured in the two groups (e.g. female vs male DI recipients), as well as the standard errors in parentheses (with significance levels displayed as *** p<0.01, ** p<0.05, * p<0.1). We use the same controls as in equation (1). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.*

Figure 11: Probability to Work while on Claim (Unconditional) for Participants who Started a Return-to-work Program in 2018 vs 2019



Notes: Horizontal axes show months since filling out baseline questionnaire and starting the return-to-work program. Vertical axis plots unconditional means for the share of study participants who work while on claim. Left panel includes only participants who entered the study between March 2018 and December 2018, while right panel includes only participants who entered the study between January 2019 and December 2019. Data source is from NIHDI administrative registers.