



Evaluation of a Placement Coaching Program for Recipients of Disability Insurance Benefits in Switzerland

Tobias Hagen¹

Published online: 9 March 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Purpose During 2009–2013 a pilot project was carried out in Zurich which aimed to increase the income of disability insurance (DI) benefit recipients in order to reduce their entitlement to DI benefits. The project consisted of placement coaching carried out by a private company that specialized in this field. It was exceptional with respect to three aspects: firstly, it did not include any formal training and/or medical aid; secondly, the coaches did not have the possibility of providing additional financial incentives or sanctioning lack of effort; and thirdly due to performance bonuses, the company not only had incentives to bring the participants into (higher paid) work, but also to keep them there for 52 weeks. This paper estimates the medium-run effects of the pilot project and assesses the net benefit from the Swiss social security system. **Methods** Different propensity score matching estimators are applied to administrative longitudinal data in order to construct suitable control groups. **Results** The estimates indicate a reduction in DI benefits and an increase in income even in the medium-run. A simple cost–benefit analysis suggests that the pilot project was a profitable investment for the social security system. **Conclusion** Given a healthy labor market, it seems possible to enhance the employment prospects of disabled persons with a relatively inexpensive intervention, which does not include any explicit investments in human capital.

Keywords Disability · Employment · Rehabilitation · Health economics

Introduction

Many countries face the problem of high costs for disability insurance (DI) benefits due to an increased number of recipients. This has led to reforms aiming to increase the employment rate of (potential) recipients [1–3]. Basically, there are two ways to achieve this: either the inflows of workers onto DI benefits can be reduced (which seems to be the most common attempt) or the outflows of recipients from DI benefits into employment can be increased [1]. As explained by Wittenburg et al. [4], employment programs for disabled workers must help them to overcome substantial employment barriers: firstly, the loss of human capital due to the

disability and the prolonged separation from the workforce; secondly, disincentives arising from the DI system, including the possible loss of entitlement to DI benefits as well as threshold effects resulting from the reduction in benefits in case of increases in incomes [5]. The latter aspect is of particular importance for Switzerland [6].

The traditional active labor market policy for this target group is vocational rehabilitation (VR), often associated with human capital investments (i.e., education and training) and/or medical aid. The idea is to help disabled people to work (again) in the profession they are trained in, or, if this is no longer possible, to retrain them by giving them the qualifications needed for new jobs [2].

In contrast to these VR measures, a pilot project was carried out in Switzerland (Zurich canton) during 2009–2013, which did not include any explicit investment in human capital or medical aid, but rather involved placement coaching by a private company specialized in this field. As a result, the pilot project was relatively inexpensive compared to training measures. The pilot project is described more detailed in the section “[The Pilot Project and the Recruitment Process](#)”. In this paper a microeconomic evaluation as well as a simple

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10926-018-9766-x>) contains supplementary material, which is available to authorized users.

✉ Tobias Hagen
tobiashagen@email.de

¹ Frankfurt University of Applied Sciences, Nibelungenplatz 1, 60318 Frankfurt am Main, Germany

cost–benefit analysis of this pilot project will be presented. Two types of propensity score matching estimators combined with difference-in-differences (DiD) are applied. The underlying *selection on observables assumption* or *conditional independence assumption* (CIA) may be rationalized by the rich administrative panel dataset used, which covers the employment history of the persons.

Empirical evidence regarding the effectiveness of job placement services without training and/or medical aid for recipients of DI benefits is scarce. Firstly, there are several studies of the effectiveness of VR measures, which include training and/or medical aid for recipients of DI benefits. Secondly, there are a lot of evaluation studies relating to job search assistance for unemployed (non-disabled) workers in general. Thirdly, there is no previous research on the effects of (intensified) job search assistance without training and/or medical aid for recipients of DI benefits. Finally, there are several studies analyzing the employment effects of in-work benefits and subsidies aiming at increasing the work incentives of recipients [7, 8]. Since financial work incentives were not included in the pilot project analyzed in this paper, this subject will not be discussed here any further.

Starting with evaluation studies on VR measures, using Propensity Score Matching Frölich et al. [9] did not find positive effects of participation in such programs compared to non-participation in Sweden. However, their results showed that workplace training was superior to the other rehabilitation measures. In contrast to this result, based on a bivariate probit model Heshmati and Engström [10] found that participation in VR programs in Sweden had positive effects on the health status as well as the rate of return to work. These contradictory results point to the fact that estimation results may be sensitive to the empirical methods applied. Recently, using several matching and weighting estimators, Campolieti et al. [11] found that a VR program in Canada improved the labor market outcomes of women, but not men. This study also attempted to provide insights into the costs and benefits of the program. From the perspective of the government, for women the expected benefits (reductions in payments) exceeded the costs. Aakvik et al. [12] evaluated VR programs for female applicants in Norway. They found a negative effect of the training program on employment prospects. The estimated effects were more positive for individuals with characteristics that predict lower employment in either the trained or untrained state. However, due to cream-skimming individuals with unfavorable employment prospects were infrequently selected into the measure. For a subgroup of workers with cognitive impairments in the state of Virginia, Dean et al. [13] were able to evaluate the effects of multiple services of VR using an instrumental variable approach. They found large positive long-run (3–9 years) effects on employment and earnings. Both the

short- and long-run mean labor market effects were estimated to be positive for diagnosis and evaluation, training, education and other services, but negative for restoration and maintenance. The mean long-run benefits exceeded the mean costs by 4–6 times. For Norway, Markussen and Røed [14] evaluated four different VR programs for temporary DI claimants. Based on longitudinal administrative data they used local variations in the policy strategies to estimate the impact of these strategies on the participants' future employment and earnings performance. Overall, they estimated positive effects. However, they found that a strategy focusing on rapid placement in the regular labor market was superior to alternative strategies giving higher priority to vocational training or sheltered employment. Summarizing the previous research on the effectiveness of VR for disabled workers, it can be said that these kinds of interventions seem to be often successful in terms of helping the participants to increase their employment prospects and in terms of reducing government spending.

There is a lot of empirical evidence regarding the effects of job search assistance and coaching for the unemployed in general. However, as summarized, inter alia, by Brown and Koettl [15] these measures may not only work because of an increase in job search and matching efficiency due to the counseling, but also because they may be associated with “threat effects” for beneficiaries, who risk sanctions in the case of a lack of effort in job search. Threat effects did not exist at all in the pilot project in Zurich since sanctions were not part of the program. This should be kept in mind when looking at the previous empirical evidence. Using a meta-analysis, Card et al. [16] found that job search assistance programs yield relatively positive effects compared to other measures such as public-sector employment programs, subsidized private sector jobs or training programs. Thomsen [17] reviewed studies related to nine European countries and concluded that job search assistance programs decrease unemployment duration and increase employment rates. This was confirmed by a review by Brown and Koettl [15]. They concluded that “measures improving labor market matching” were cost-effective and might have significant short-run effects. They also concluded that lock-in effects might have been minimal, which seems to be important for the understanding of this paper's results. The so-called “lock-in effect” refers to the initially lower probability of finding a job of participants in active labor market programs compared to the non-participants, which may occur because participants spend less time and effort on job search activities than non-participants [18]. Furthermore, Brown and Koettl [15] concluded that the programs should be targeted at persons with poor employment prospects at the beginning of their unemployment spell and at long-term unemployed persons. Moreover, they found that these programs were most effective during economic upswings.

Finally, there is only one empirical study evaluating a program with some similarities to the pilot project in Zurich. Høgelund and Holm [19] evaluated the effect of case management interviews (CMIs) performed by social caseworkers on the probability of a return to work by disabled employees in Denmark. Based on instrumental variables and a competing hazard rate model, they found that CMIs increase the probability of returning to work for the pre-sick leave employer, but have no effect on the probability of commencing work for a new employer. However, as the CMIs are made by the municipal case managers they have the availability of all VR instruments that may help the DI benefit recipients. This is definitely not the case in the evaluated pilot project in this paper.

This paper presents an econometric evaluation of a placement coaching program for DI beneficiaries. Placement coaching programs may be of special interest for policy-makers, as they are relatively inexpensive in comparison to formal training programs [15]. On the one hand, the (almost complete) lack of investment in human capital may hamper the effectiveness of such a program at least in the long-run. On the other hand, it may increase the probability that the benefits of the program (a reduction in DI payments) will exceed these costs. The estimates in this paper point to a successful project in terms of a reduction in DI benefits and an increase in income even in the medium-run (4 years after the program start). The cost–benefit analysis indicates that the project was a profitable investment for the social security system: depending on the scenario assumed, the expected mean long-run benefits exceed the mean costs by 1.9–6.5 times.

The remainder of the paper is organized as follows: The Swiss DI system is briefly outlined in the section “[Institutional Background: Swiss DI](#)”. The section “[The Pilot Project and the Recruitment Process](#)” describes the pilot project as well as the selection mechanism for it. The former is important for the understanding of possible causal mechanisms for outcome variables as well as the incentives of the private company carrying out the project. The latter is useful for the specification of the propensity score equation. The administrative data used and the sample are described in the “[Data and Sample Selection](#)” section. The “[Econometric Approach](#)” section presents the econometric approach (propensity score matching) and its application to the subject of the research. The empirical results on the propensity score, the “match quality” and the estimated effects are shown in the “[Empirical Results](#)” section. Sensitivity analyses in the Web Appendix provide additional information on the robustness of the estimated effects. The “[Assessment of Costs and Benefits](#)” section assesses the resulting costs and benefits of the pilot project from the social security system’s perspective. The paper concludes with a summary and a discussion of the policy implications.

Institutional Background: Swiss DI

The Swiss social security system in general, together with the Swiss DI in particular, is based on a “three pillar system”. The first pillar is a state pension plan, including the DI. The second pillar consists of occupational pension plans and accident insurance, which are mandatory for (almost) all employees. The third pillar is employees’ private provision, which should complement the first two pillars. The third pillar is also protected by law and is often promoted by tax facilities. The pilot project and all administrative data are related to the first pillar. This implies that information about the second and the third pillars is not available.

The DI, as a part of the first pillar, aims to guarantee the basic needs of insured persons who have become disabled, by paying DI benefits and/or by providing rehabilitation measures.¹ Disability is defined as a decline in the ability to earn a living or in the ability to accomplish daily tasks, such as housework, resulting from physical, psychological or mental health problems. To qualify as a disability this incapacity must last at least 1 year. When judging, if an inability to earn a living is present, it does not matter what the causes of the health depreciation are.

Insured persons who have paid contributions for at least 3 years can claim a DI benefit. The right to a DI benefit begins when the insured person has an average incapacity to work of at least 40% (the so-called *degree of disability*) and after a waiting period of 1 year following the end of employment for health reasons. The degree of disability corresponds to the percentage loss in earnings relative to the potential earnings without the disability [6]. There are four different *DI benefits entitlement* types, depending on the degree of disability. These range from 25 to 100% (Table 1).

Moreover, Table 1 provides a descriptive statistic on the sample used here. It can be seen that the participants are a positive selection with regard to the degree of disability: the mean DI benefit entitlement of the participants is 72.6%, versus 81.9% in case of the non-participants. Nevertheless, one should keep in mind that more than half (54.6%) of all participants are “full pensioners”. That means, their degree of disability is at least 70% and they receive a full (100%) DI benefit.

Traditionally, job placement and vocational guidance take place before persons are qualified for DI benefits. As described in the next section, the pilot project addressed this issue by focusing on DI benefit recipients. The participants’ average duration of DI benefit receipt was more than 5 years (see Table 2 in the Web Appendix).

¹ This section is based on <http://www.zas.admin.ch/org/00858/00861/index.html?lang=en>.

Table 1 Degree of disability and the type of DI benefit

DI system		Proportions in the sample in%	
Degree of disability	DI benefit entitlement	Non-participants (potential control observations)	Participants in the year before the individual program start
< 40%	No benefit	5.8	9.6
≥ 40 < 50%	25% benefit	4.4	7.2
≥ 50 < 60%	50% benefit	14.9	21.2
≥ 60 < 70%	75% benefit	6.3	7.5
≥ 70%	100% benefit	68.7	54.6
Average DI benefit entitlement in %		81.9	72.6
Number of observations		40,710	908

The descriptive statistics are based on the estimation sample described in “[Data and Sample Selection](#)” section

The Pilot Project and the Recruitment Process

To implement the pilot project an international private company that focuses on workforce participation was commissioned. The project consisted mainly of placement coaching by individually assigned advisers/coaches. Figure 1 provides an overview of the project. During the entire placement process, the participants received active support in, and practical tips on, their search for suitable jobs. In addition to providing assistance in preparing job applications, the coaches discussed career prospects with the participants, searched for potential positions together with them, and provided them with the materials and postage needed for applications. Supplementary courses (often lasting for only a few hours) were offered on topics such as self-management or job application techniques, but never in the sense of vocational training. The coaches used publicly available job advertisements. The company did not have its own vacancy database. However, due to the excellent labor market conditions in Zurich (see Footnote 5), this was a minor issue. The coaches did not have the possibility of providing any financial incentives, or to sanction their clients for a lack of effort. The practical help did not only contain job search and vocational guidance, but all kinds of consulting, which might have helped to raise the participants’ employability. For example, this could include talks (but not in sense of psychological treatments) about the organization of everyday life or living with disabilities. The company stressed the importance of participants developing self-esteem and motivation. One approach to achieve this, was the company’s endeavor that the participants felt as a part of a team, instead of being alone at home. This idea was reflected in the premises in which the project took place: it was an open-plan office, which should promote mutual learning and motivation. The coaches had relatively high formal qualifications since a university degree was the precondition for being recruited by

the company. Newly recruited coaches were trained by the company in relevant coaching techniques and the coaching guidelines of the company.

The “placement phase” lasted for a maximum of 12 months. Those who dropped out prematurely were given the option of starting the program again. “Dropouts” are participants who drop out within the first 3 months of the program, because they did not show up at appointments with their coaches. Hence, some of the dropouts may be individuals who found jobs successfully on their own, but who did not want to cooperate with their advisers anymore. In total, 151 (16.6%) out of 908 participants dropped out. Five (3.3%) of the dropouts restarted the treatment later.

Participants who were successful in finding jobs received follow-up support from the company for up to 12 months in order to stabilize the employment relationship (“follow-up phase”). This support consisted of coaching again. Those individuals who subsequently quit their jobs or were dismissed were not excluded from the measure, but could participate again. The placement company was paid “sustainability bonuses” depending on the length of the employment relationship achieved (26 or 52 weeks). However, the wage had to amount at least to 50% of the customary wage in the respective industry.

The DI benefits were not reduced until the participants had completed their probationary period in the new job (at the earliest after 3 months). Note that this is not a special feature of the project, but it is rather the regulated process for all DI benefit recipients.

In addition to the sustainability bonuses of CHF.² 3000 paid for every participant placed in a new job for a period of 26 weeks (or the double amount for 52 weeks), the invested amount in the pilot project by the DI scheme comprised a lump sum per case of CHF 6000 and overall

² CHF=Swiss Franc, currency and legal tender of Switzerland.

set-up costs of CHF 2.28 million. The lump sum of CHF 6000 per case was paid after 3 months of participation. The total cost per participant was CHF 8819.

In summary, the pilot project was exceptional with respect to three aspects: firstly, it did not include any formal training and/or medical aid; secondly, the coaches/advisors did not have the possibility of providing additional financial incentives or making use of “threat effects”; and thirdly, due to the bonuses involved, the company had financial incentives not only to secure the participants in (higher paid) jobs, but also to keep them employed for 52 weeks.

Before DI benefit recipients could participate in the pilot project, they had to complete the three-stage process depicted in the far-left of Fig. 1: in the *first stage*, the DI office in Zurich recruited potential participants from the population of almost 50,000 DI benefit recipients in Zurich canton. They were informed of the possibility of participating and the implications for their entitlement. People aged between 18 and 58 were targeted, the intention being to achieve an age distribution that reflected that of the entire population of DI benefit recipients. The participants needed to exhibit reintegration potential. At the very least, there had to be reasonable grounds for assuming that they could achieve reintegration potential, which was determined by the administrative staff of the DI office on subjective assessment. In addition, those DI benefit recipients were considered who said that their state of health had improved. Insured persons presumed to be incapable of carrying out paid employment were not actively recruited, nor were those who had never worked before. However, in individual cases, the latter were allowed to take part on their own initiative. In this manner, a total of more than 15,000 persons were recruited for the project. This group forms the basis for the sample of the participants as well as the sample of the control group generated by the matching estimators. Put differently, the almost 35,000 DI benefit recipients in Zurich, who were not recruited, are not used as potential controls.

In the *second stage*, 1368 people interested in taking part (participation was not mandatory) received a ruling from the Zurich DI office, which means that they got an official document granting permission to participate.

In the *third stage*, those persons who had received a ruling were invited for a preliminary talk with the placement company. As some of the invitees either did not respond or decided after the preliminary talk not to take part, not every ruling resulted in participation. A total of 947 persons took part in the project between November 2009 and August 2011. For reasons relating to methodology, as explained in the next section, the evaluation is based on 908 participants only.

Table 2 Number of individuals: treated and potential controls in December 2009

All DI benefit recipients in Zurich	49,951
Persons recruited ^a	14,878
Persons who received rulings ^a	1037
Participants (2009–2011) ^a	908

^aIndividuals who died until December 2014 and people who transitioned to the old-age pension system are excluded

Data and Sample Selection

The whole analysis is based on administrative data that the Federal Social Insurance Office has gathered from six data registers. For the period 2000–2014 these data include information on DI benefits and further wage-replacement payments, diseases, socio-economic characteristics (age, nationality, gender etc.), participations in rehabilitation measures and income. The variables are discussed in greater detail below.

The number of individuals is shown in Table 2. The total number of recipients was almost 50,000. Out of these 14,878 individuals were recruited for the pilot project by the DI office in Zurich. However, in this number as well as in the other subgroups marked with an asterisk, the people who died until 2014 are already excluded. This represents 525 (3.4%) of all the recruited individuals. For the subgroup of participants this amounts to 19 persons (2%). Furthermore, individuals who transitioned to the old-age pension system before December 2014 are excluded. This is the case for 24 (0.16%) of the recruited persons and for eight (0.87%) participants. The removal of these individuals from the sample is based on the fact that the outcome variables are not observable for old-age pensioners and dead persons. The deletion of dead persons is definitely not a problem with regard to selection bias as long as participation does not affect mortality. In contrast, the removal of those persons who transitioned to the old-age pension system could lead to a positive selection of the sample. However, this is relevant for less than 1% of the treated individuals and since also the corresponding non-participants are dropped, this kind of sample selection bias should be a minor issue.

The data on DI benefit entitlements and payments are measured each year in December for the period 2000–2014, and therefore the whole analysis can be based only on an annual frequency. The potential control group consists of individuals who were recruited but did not participate (14,878 persons in Table 2). Given this approach and after eliminating non-participants with missing data (either outcome or conditioning variables), Table 3 shows the number of participants and potential controls available each year.

The outcome variables are shown in Table 4. The outcome variables measured in CHF are adjusted either for DI

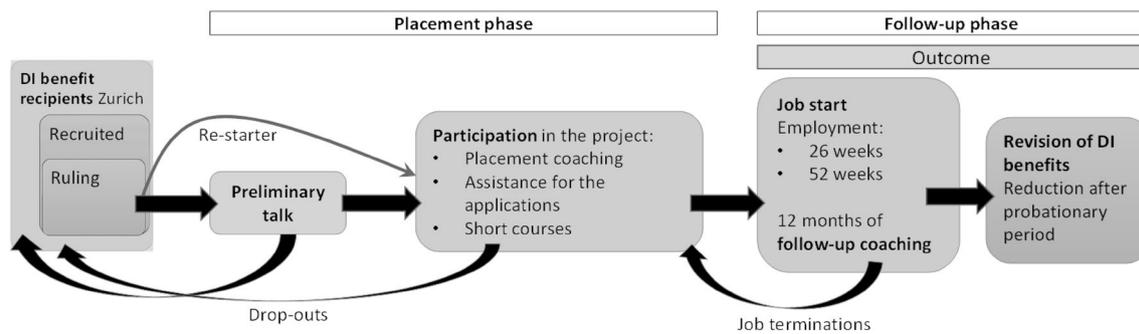


Fig. 1 Flowchart of the pilot project

Table 3 Number of program-starters and non-participants by year

Year	Participants (program starters)	Nonparticipants (potential controls)	Total
2009	52	13,570	13,622
	0.38%	99.62%	100%
2010	527	13,570	14,097
	3.74%	96.26%	100%
2011	329	13,570	13,899
	2.37%	97.63%	100%
Total	908	40,710	41,618
	2.18%	97.82%	100%

benefit increases or consumer price inflation to the base year 2009.

The following conditioning variables are available for the analysis (see Table 2 in the Web Appendix for sample means):

- Socio-demographic characteristics: standard variables such as age, gender, nationality, civil status and number

of children are available. Variables measuring the educational background are not available.

- Health status: no subjective information is available with regard to well-being. However, all the information that is necessary for the application and the approval of DI benefit is contained. In concrete terms: with the type of disease it is possible to distinguish between congenital, mental, nervous system, injuries, and other diseases. For the participants' mental diseases (55.7%) are most important, followed by musculo-skeletal defects (15.1%). The variable functional disorder indicates the implications of the diseases for the individual employment prospects. For the analysis the five most frequent functional disorders are used and the others are pooled into one residual category. Furthermore, the data include information on the so-called helplessness allowance, to which people who permanently require a considerable degree of help from a third person are entitled.
- Occupational history: total income subject to deduction of social insurance contributions and income from paid employment are available. Previous participation in VR measures under the DI system can be identified. The

Table 4 Outcome variables

Outcome variable	Period	Adjusted for	Notes
Monthly main DI benefit in CHF	December 2000–December 2014	DI benefit increases, every second year	Measured in December each year
Monthly total DI benefit in CHF (main benefit and child's benefit)	December 2000–December 2014	DI benefit increases, every second year	Measured in December each year
DI benefit entitlement in %	December 2000–December 2014		Measured in December each year; see Table 1 for an explanation
Monthly supplementary benefits (SB) per case in CHF	December 2000–December 2014	Consumer price index, annual	Total payment for households
DI benefit recipient (yes) in %	December 2000–December 2014		Dummy variable
Annual income earned from paid employment in CHF	2000–2013	Consumer price index, annual	Calculated from contribution to social security
Income earned from paid employment (yes) in %	2000–2013		Dummy variable

daily allowance of the DI system also indicates that a person participated in a VR measure. The receipt of an *allowance of the unemployment insurance system* indicates that a person has paid sufficient contributions and is still in the labor force.

- DI benefits: the amount of *monthly DI benefit* (so-called “main” and “child” benefit) for every December since 2000 is included in the data. *Extraordinary benefits* are for people who became disabled before their 20th birthday.
- Additional social benefits: if the income (including DI benefits and other income sources) does not suffice for living, the DI benefit recipients may claim *supplementary benefit*.

Three possibly important groups of variables are missing: information on educational background, personality traits, and further income sources (the second and the third “pillar” of the Swiss system). This is an obvious methodological shortcoming of this paper, because confounding variables may affect incentives and, hence, the selection into the program, as well as outcome variables. Put differently, the fact that these possibly relevant variables are not observed may lead to selection bias. The next section will discuss how to deal methodologically with this problem of missing variables. Due to its long time dimension, the dataset seems nevertheless to be relatively comprehensive. This becomes obvious when comparing it to, for example, the well-known dataset used by LaLonde [20], Dehejia and Wahba [21], Smith and Todd [22] and others.

Econometric Approach

Basics

The goal is to estimate the average treatment on the treated (ATT) effect of the coaching program on a set of future outcome variables, namely *income*, *DI benefits* and *supplementary benefits* (see Table 4). The evaluated treatment is the start of the measure, which implies that “dropouts” are included. The counterfactual is defined as “complete non-participation in the program”. The non-participants are defined as individuals who were selected by the DI office as potential participants (were recruited), but decided not to participate (see Fig. 1; Table 2). They had the possibility of participating in another public-provided program. The latter is a reasonable definition since the evaluated treatment is a temporary pilot project, which was additional to the existing VR programs.

In recent years, following Fredriksson and Johansson [23] as well as Sianesi [24, 25] the static evaluation approach to training measures has been criticized for leading to biased

estimates: if, based on the erroneous assumption that a program is administrated only once, the control group is defined as non-participants who never participate, the researcher conditions on future outcome variables [26]. This is definitely not a problem here: *firstly*, the pilot project was in fact administered only once. All potential participants were informed before the start of the whole measure. *Secondly*, there was no excess demand and hence no queue of persons waiting for an opportunity to participate. On contrary, it was hard to fill all available positions: out of 14,878 informed (recruited) persons, only 947 (6.4%) participated. *Thirdly*, as the program was rather small in scale, there is a relatively large number of non-participants (approx. 13,500 per year) that can be used as a control group for the 908 participants. Against this background, the standard static approach can be safely applied. This means, that non-participants can be defined as individuals who never participate in that program, but do possibly participate in another traditional VR measure.

The econometric approach used to estimate the ATT effect is Propensity Score Matching.³ Since this method has become standard in applied research, it is briefly summarized only. The identifying assumption is called the *conditional independence assumption* (CIA) or *unconfoundedness* assumption [28]. For the ATT the CIA is $Y_0 \perp C | X$, denoting that the outcome of the non-treated individuals (Y_0) and the participation (C) are independent conditional on the observable variables X . The second assumption is *common support* or *overlap* denoted as $\Pr(C = 1 | X) < 1$ for the ATT, that means, that every participant has a positive probability of being non-participant. Matching on the propensity score is based on the “modified” CIA $Y_0 \perp C | e(X)$, with $e(X) = \Pr(C = 1 | X)$.

The estimation procedure is as follows: the propensity score equation for the participation is estimated with a probit model based on $N_1 = 908$ participants and up to $N_0 = 40,710$ observations of non-participants and the observed conditioning variables X . X includes pre-treatment outcome variables. Afterwards the individual propensity score $\hat{e}(X) = \Pr(C = 1 | X)$ for all $i = 1, \dots, N_1$ treated $\hat{e}_1(X)$ and $j = 1, \dots, N_0$ and untreated individuals $\hat{e}_0(X)$ is predicted.

After matching, the ATT effect is calculated as

$$\frac{1}{N_1} \sum_{i=1}^{N_1} \left(Y_{1i} - \sum_{j=1}^{N_0} w(i,j) Y_{0j} \right)$$

³ Here “psmatch2” implemented in STATA by Leuven and Sianesi [27] is applied. STATA is the statistical software created by StataCorp LLC, 4905 Lakeway Drive, College Station, Texas 77845-4512, USA.

with Y_{1i} indicating the outcomes of the treated individuals and Y_{0j} being the outcomes of the non-treated individuals. To every non-treated (within common support) the weight $w(i, j)$ is attached, with

$$\sum_{j=1}^{N_0} w(i, j) = 1$$

and $w(i, j)$ being a negative function of the distance, in terms of $\hat{e}(X)$ (or X directly), between the treated individual and the corresponding control individuals.

In case of kernel-based matching the weight $w(i, j)$ is calculated as

$$w(i, j) = \frac{G\left(\frac{e_j - e_i}{h}\right)}{\sum_{k \in (C=0)} G\left(\frac{e_k - e_i}{h}\right)}$$

where $G(\cdot)$ is the kernel function and h is the bandwidth parameter. The Epanechnikov kernel is applied here, which has the advantage of attaching a weight of zero to control observations outside the bandwidth [29].

In recent years there has been an increasing insight into optimal bandwidth choice in kernel matching [26, 29]. Silverman’s [30] rule of thumb suggests a bandwidth value of 0.06, which will be used here. Surprisingly, and in contrast to the literature, the estimated ATT effects here are relatively robust to changes in the bandwidth (see Section III.2 in the Web Appendix).

The standard errors for the estimated ATT effect are obtained by a bootstrapping procedure over both steps (propensity score and matching) with 250 resamples.

Recently, Lechner et al. [31] proposed a distance-weighted radius matching with bias adjustment.⁴ Here, this new approach is applied as a kind of sensitivity analysis, in addition to the kernel-based matching estimator. The basic idea is “caliper matching”, extended with a bias-adjustment based on linear regressions. Huber et al. [32] found that including the most important covariates (on top of the propensity score) in the matching algorithm (via the Mahalanobis metric) lead to better results in terms of decreasing the selection bias. Here the variables *gender*, *calendar year* (2009, 2010, 2011) and *DI benefit entitlement in %* before the treatment are included. This guarantees that controls must have the same gender and the same DI benefit entitlement in % as their corresponding treated individuals. Furthermore, control observations come from the same calendar year as treated persons. With respect to the tuning

parameters discussed in Huber et al. [32], the default values suggested by the authors are chosen. Due to the computing time it is not feasible here to bootstrap the standard errors and therefore inference must be based on analytical standard errors.

In order to eliminate the possible bias due to selection on unobservables, the matching procedure is extended by DiD [22]. The ATT effect is estimated as

$$\frac{1}{N_1} \sum_{i=1}^{N_1} \left((Y_{1i\tau} - Y_{1i,t-1}) - \sum_{j=1}^{N_0} w(i, j) (Y_{0j\tau} - Y_{0j,t-1}) \right)$$

with $\tau = \{t + 1, t + 2, t + 3, t + 4\}$ and t being the year of the individual program start (2009, 2010 or 2011). In contrast to the calculation of the ATT effect as difference in post-program “levels” of outcome variables ($Y_1 - Y_0$), the ATT effect is now calculated as the differences of changes of Y_1 and Y_0 over time. Smith and Todd [22] found that DiD matching estimators exhibit better performance than cross-sectional matching estimators, because time-constant (selection) biases are differenced out by this method. This may be helpful here, since no information on educational background, personality traits, and further income sources (the second and the third “pillars” of the Swiss system) are available. Caliendo et al. [33] showed that conditioning on individuals’ labor market histories in administrative datasets may help to reduce the bias from unobserved variables such as personality traits, attitudes, expectations, and job search behavior (see also [34]). With regard to these variables, risk attitude may be especially relevant here: if a DI benefit recipient takes up a sufficiently well-paid job, she may lose her entitlement to DI completely. This implies that this person must go through the whole application process again, if she becomes unable to earn a living again in the future. Sensitivity analyses on the potential bias due to selection on unobservables/confounders in the Web Appendix indicate that this may be a minor problem.

DiD and conditioning on pre-treatment outcome variables in general require that there are no “anticipatory effects”. For example, an anticipatory effect may mean that future participants in retraining measures for unemployed workers reduce their search effort before the start of the treatment, because unemployment is an eligibility criterion for participation [35, 36]. Hence, the outcome before the treatment $Y_{1,t-1}$ would be affected by the treatment and the DiD estimator would be biased. Below it will be argued that anticipatory effects are unlikely here.

⁴ Huber et al. [32] implement this estimator in STATA with the command “radiusmatch”.

Application of Propensity Score Matching to the Evaluation Problem Here

Due to issues relating to data availability all matching procedures are based on two samples. The reason for this is that the outcome variables for DI benefits are available until December of 2014 and the outcome variables for incomes are available until the end of 2013. The individual program starting years, t , are between 2009 and 2011. This implies the following two samples:

Sample “*Program starters 2009–2011*”: the ATT effects on the DI benefit outcome variables for $t + 1$, $t + 2$ and $t + 3$, as well as the effects on the employment outcome variables for $t + 1$ and $t + 2$, can be estimated for the participants who started in 2009, 2010 and 2011. This sample includes 908 individuals.

Sample “*Program starters 2009–2010*”: the ATT effects on the DI benefit outcome variables for $t + 4$, as well as the effects on the employment outcome variables for $t + 3$, can be estimated for the participants who started in 2009 and 2010. This sample includes 579 individuals (= 52 starters from 2009 + 527 starters from 2010).

As mentioned in the previous section, the propensity score is estimated by a probit model. This is done separately for the two samples. The specification with regard to the conditioning variables does not follow theoretical arguments [28], but it is based on the following considerations. On the one hand, it seems plausible to include all available variables, including the pre-treatment outcome variables, in the conditioning set since all variables are potential determinants for outcome as well as selection into treatment. The pre-treatment outcome variables may serve as proxies for missing variables. For example, the previous income is likely to be highly correlated with the (unobserved) educational background. Previous research also suggests that higher order terms of variables and/or interaction terms of variables could be included in order to achieve a “balanced” control group with respect to X [21, 37]. Moreover, as panel data are available it is possible to include lagged values of time-varying X (including Y) up to $t - 7$. On the other hand, multicollinearity problems and a possible increased variance of the estimates or even inconsistent estimates due to “too many” covariates in the probit model suggest a parsimonious specification [28, 33, 38, 39].

Against this background, the search for an “optimal” specification of the propensity score equation is guided by the following two criteria. Firstly, those specifications are preferred which balance pre-treatment outcome variables up to $t - 7$. For this criterion the so-called “pre-program test” [35] is shown in the section “*Balancing of Covariates*” for the preferred specification. Secondly, the propensity score matching should balance the other pre-treatment conditioning variables in X . Based on these criteria the preferred

specification of the propensity score probit is found. In particular, *age* is the only variable that is included squared. For some pre-treatment outcome variables (*monthly main DI benefits, incomes*) and further conditioning variables lagged values up to $t - 3$ are included. Variables are included even when they do not have a statistically significant effect in the probit. The results of the propensity score estimate will be presented in the next section.

As explained in the previous “*Basics*” section, the inclusion of lagged outcome variables and the DiD estimator are based on the assumption that there are no “anticipatory effects”. This assumption seems plausible here because the (potential) participants had no incentive to change their behavior prior to the start of the measure. No direct financial advantage or disadvantage arises from the participation. Also, the empirical data in the section “*Balancing of Covariates*” do not show a “dip” in $t - 1$ or $t - 2$ [35, 40]. Given this, it seems valid to condition on pre-treatment outcome variables and to apply DiD.

Another issue is how to deal with “time” in terms of the individual starting year, $t = \{2009, 2010, 2011\}$, of the measure and the calendar year of the non-participants. There are three possible approaches, as set out in the following paragraphs.

Firstly, strictly define that controls must have the same t as their corresponding treated individuals. This approach seems necessary, if the reemployment opportunities of the DI benefit recipients change over time. In international comparison, the labor market conditions in Zurich are excellent. From 2009 to 2014 there had been some variations in the overall unemployment rate in Zurich.⁵ However, it is unclear whether and to what extent this is relevant for the DI benefit recipients. In case of radius matching, t is included into the Mahalanobis distance (in addition to the propensity score). With this approach it is (almost) guaranteed that the strict definition is fulfilled. A potential drawback of this strict definition is the resulting reduction in the number of potential control observations to 13,570 per year (Table 3).

Secondly, ignore t and allow that all 40,710 non-participant observations are potential controls. This would be a valid approach, if changes in the labor market situation over time were not relevant.

Thirdly, define the t as a “weak restriction” in the sense that it is only included in the explanatory variables of the propensity score. Hence, t is one conditioning variable, alongside others included in X . This is the approach chosen here for the kernel-based matching since it leads to the best results in terms of balancing the other covariates and the

⁵ Average annual unemployment rates: 3.7% in 2009, 3.6% in 2010, 2.9% in 2011, 3.0% in 2012, 3.2% in 2013, 3.3% in 2014. Source: own calculation based on <http://www.amstat.ch>.

pre-treatment outcome variables. However, it comes with the cost that t is not identical for all treated and corresponding controls. Note that this approach means that the same non-treated person can be used three times as control with different weights.⁶

Without changing anything explained above, matching is here not on the propensity score, but on the underlying linear index [41]. After the estimation and prediction of the propensity score the common support condition is examined. For every participant it is checked whether the estimated propensity score (linear index) is overlapped by the estimated propensity scores of untreated individuals.

Empirical Results

Propensity Score

Due to high collinearity of the coefficients the propensity score estimates have no causal interpretation. The estimation results can be found in Section I in the Web Appendix.

After the estimation of the propensity score the common support condition is analyzed. A graphical representation for both samples can be found in Figure 5 in the “Appendix”. For the “program starters 2009–2011” sample the propensity scores of all treated individuals are overlapped by the scores of untreated individuals. In the upper tail of the density this is less obvious. However, for all treated individuals the estimated propensity score is lower than the maximum propensity score of the controls. In order to clarify the large number of potential controls, histograms with the absolute number of observations are shown in the lower graphs. Hence, in the “program starters 2009–2011” sample common support is given for all $N_1=908$ participants.

The graphs for the “program starters 2009–2010”-sample show a bimodal distribution for the untreated individuals. This phenomenon is generated by the two starting years.⁷ Again the treated observations seem to be overlapped. However, support is not given for one treated person (its propensity score is higher than the maximum propensity score of the untreated sample) and thus the estimates are based on 578 treated individuals only.

⁶ Although this may seem odd at first glance, using the same person as a control for many times is a common practice, for example, in nearest neighbor matching with replacement. Also a pooled panel regression can be interpreted as using untreated individuals several times as controls.

⁷ For example, the median propensity score (linear index) of the untreated individuals is -2.77 in 2009 and -1.88 in 2010.

Match-Quality

Balancing of Covariates

Balancing tests are based on the property $XC|\hat{e}(X)$: after matching on the propensity score (and possibly further conditioning variables) the treatment status, C , should be independent from the conditioning variables, X . Put differently, there should not be significant differences between treated and controls with respect to the conditioning variables, X . There is not one sole balancing test. Different approaches have been proposed. Lee [42] provided an overview.

Table 2 in the Web Appendix shows a detailed analysis separately by every covariate included in the propensity score. It shows balancing tests with respect to each covariate for both periods (program starters 2009–2011 vs. 2009–2010) as well as the unmatched (U) and the matched (M) samples. First of all, the means of each variable in the treated group (\bar{x}_1) and the untreated group (\bar{x}_0) are presented. The differences between them are much smaller in the matched samples than in the unmatched samples. This is confirmed by a t test with the null hypothesis that the difference is zero. While the differences ($\bar{x}_1 - \bar{x}_0$) are large and often statistically significant in the unmatched samples, the differences become small and are always insignificant at the 10% level in the matched samples. This is confirmed by the standardized differences (std. diff. %, see the notes to Table 2 in the Web Appendix) being significantly reduced in the matched samples in comparison to the unmatched samples. This is also the main insight from Fig. 2, which presents the standardized differences of all condition variables in histograms. There are two exceptions to this statement for the time period 2009–2011: the imbalances in the dummy variable *year 2011* and in the variable *number of child DI benefits* slightly increase. However, the differences are still statistically insignificant. With regard to the standardized differences in the matched samples the question arises whether they are “small enough”. Rosenbaum and Rubin [43] designated a standardized difference of greater than 20% as large. Caliendo and Kopeinig [33] argued that in most empirical studies standardized differences below 3 or 5% are seen as sufficient. Here, all standardized differences are considerably smaller than 20%, and most are even smaller than 3%.

Finally, Table 2 in the Web Appendix shows a measure proposed by Rubin [44]. The ratio of the variance of the residuals orthogonal to the linear index of the propensity score in the treated group over the untreated group is calculated for each variable. If a variable is perfectly “balanced” this variance ratio is 1.0. Again, it can be seen that the measure is close to 1.0 for all variables in the matched samples. Moreover, for almost all variables the variance ratio is closer to 1.0 in the matched sample than in the unmatched sample.

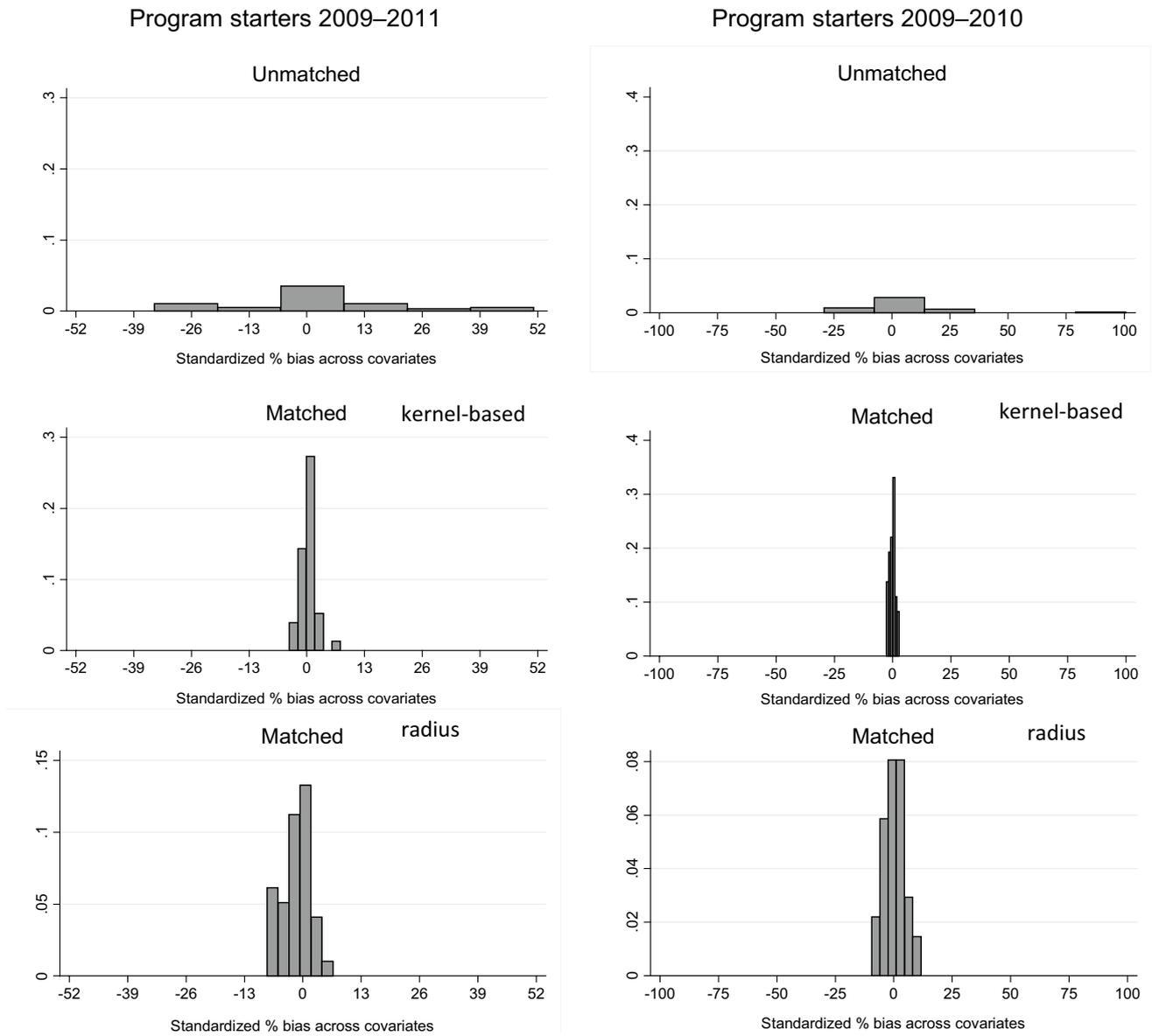


Fig. 2 Histogram of standardized differences of all conditioning variables

Table 5 Match quality, summary measures

Matching method		Pseudo-R ²	p value Likelihood ratio test	Mean bias	Med. bias	Rubin's B	Rubin's R
Program starters 2009–2011							
–	U	0.140	0.000	14.1	7.9	116.7	1.21
Kernel-based	M	0.003	1.000	1.3	0.9	12.3	0.99
Radius	M	0.007	0.999	2.7	1.8	19.5	0.84
Program starters 2009–2010							
–	U	0.127	0.000	12.4	6.4	118.8	0.72
Kernel-based	M	0.001	1.000	1.1	0.9	7.8	0.91
Radius	M	0.011	0.999	3.6	3.0	22.4	1.24

The next step is to look at summary measures of the overall (im)balance of all conditioning variables. These are shown in Table 5 for kernel-based matching as well as for radius matching. The Pseudo- R^2 is from a probit estimate of the propensity score equation in the unmatched and the matched samples. The fact that the Pseudo- R^2 is near zero in the matched samples indicates that after matching the conditioning variables no longer have any predictive power for the participation. This is a further indication that differences between treated and control individuals are balanced. The p value of the likelihood ratio test of the joint significance of all explanatory variables in the probit model points in the same direction. The following two columns in Table 5 show the mean and median of the absolute value of standardized differences of all variables. For example, due to the kernel-based matching procedure the mean standardized difference is reduced from 14.1 to 1.3 for the period 2009–2011. In both of the final columns of Table 5 two summary measures proposed by Rubin [44] are shown. Rubin's B is the absolute standardized difference of the means of the linear index of the propensity score in the treated and the (matched) untreated group. Rubin [44] specified that a B below 25 indicates a balanced control group. With the values 12.3 and 7.8 this is given for both matching estimators. Rubin's R is the ratio of treated to (matched) untreated variances of the propensity score index. This latter measure should be between 0.5 and 2. This is again the case in the matched samples of both matching estimators.

Given these analyses, one can conclude that both matching estimators are able to balance the pre-treatment differences between the treated and the control group. Hence, the specification of the propensity score equation seems sufficient in the sense that no additional polynomial terms or interactions of variables are needed.

However, important for the following analyses is the result that the kernel-based matching estimator performs better than the radius matching estimator in terms of balancing the conditioning variables. This is true for all summary measures in Table 5 as well as in the histogram of the standardized differences in Fig. 2. For this reason, the kernel-based matching estimates are seen as the “preferred specification”. The results of the radius matching estimator are interpreted as robustness checks.

Pre-program Outcome Variables

The “pre-program test” is based on the consideration that, if no selection bias remains, there should be no significant differences between the mean outcome variables of treated individuals and the control individuals before the individual start of the participation. Put differently, the ATT effects before the treatment should be zero.

The following analyses are shown for the kernel-based matching estimator only, which is defined here as the “preferred specification”. Figure 3 presents the evolution of mean outcome variables by group (treated, untreated, controls) over time as well as the corresponding 95% confidence intervals (CI). Here, time means years before (–7 to –1) and years after (+1 up to +4) the year of the individual starting year, $t = \{2009, 2010, 2011\}$. The graph on the top left shows the average monthly DI benefit in CHF by group. The increase in the mean benefit up to the year of individual program start (0) results from the fact that non-recipients are included with a benefit of zero. At the end of the individual starting year of the program, almost all participants are recipients (see the middle left graph). When comparing the untreated with treated persons it becomes obvious that the treated group receives lower benefits on average. When comparing the treated with the control group after matching before the start of the program (–7 to –1) no statistically significant differences between treated and controls remain. Hence, the outcome variable *monthly DI benefit* passes the “pre-program test”. The same is true when looking at *total DI benefit* (main + child benefit) in the top right graph in Fig. 3. Furthermore, both graphs in the top row already indicate that the program is effective with respect to the reduction of DI benefits: while after the start of the program (+1, +2...) the benefits of the control group stay almost constant, there is a significant reduction in the treated group. The other graphs in Fig. 3 indicate why the program seems to be effective: the bottom left graph shows the participation in the program increases income from paid employment, while the income of the control group is constant or may even decrease. The graphs in the middle indicate that this increase in earnings reduces the number of recipients as well as DI benefit entitlement in %.

Estimated ATT-Effects

Table 6 presents the estimated ATT-effects on DI benefits and supplementary benefits. The effects are estimated by kernel-based matching (“kernel”) as well as radius-matching with bias adjustment (“radius”). For kernel-based matching both the ATT effects estimated in levels and the ATT effects estimated as DiD are shown. The standard errors of the kernel-based matching results are bootstrapped (250 replications). Table 7 presents the corresponding ATT effects on income variables. As explained in the section “Data and Sample Selection” income is observed only up to 2013 and hence ATT effects cannot be estimated for $t + 4$.

The results of the three methods differ only slightly, which may be an indication of robustness. Focusing on the “kernel, DiD” estimates, one may draw the following conclusions: all the findings indicate that, on average, the program is effective for the participants in that the DI

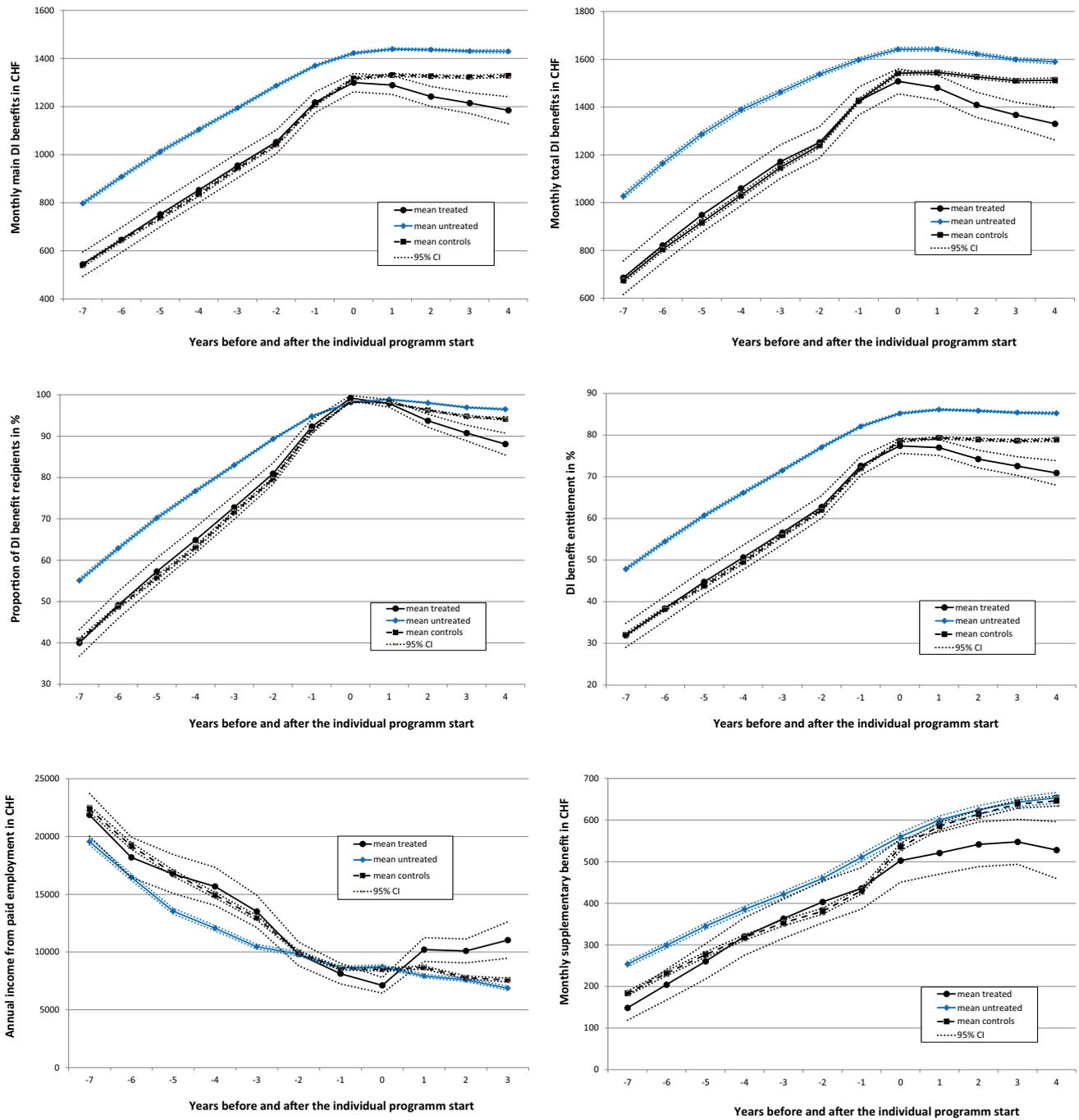


Fig. 3 Mean outcome variables of the treated, untreated and control group. All variables are prices adjusted to the base year 2009. Persons with no DI benefits are included with a benefit of zero. Also, persons

with no income from paid employment or supplementary benefit are included with a value of zero. See Table 4 for definition of the outcome variables

amounts paid to them are lowered by raising their levels of paid employment. Compared with the control group, the proportion of DI recipients among the participants is 3.4% points lower in the second year after individual starting

year ($t + 2$), 4.9% point in $t + 3$ and 7.8% points in $t + 3$ (Table 6). In the group of participants in $t + 2$, the proportion of persons earning income from paid employment is

Table 6 Estimated ATT-effects on DI benefits and SB, *t* stats in parentheses

Outcome variable	Matching method	<i>t</i> + 1	<i>t</i> + 2	<i>t</i> + 3	<i>t</i> + 4
		Starters 2009–2011 Outcome 2010–2012	Starters 2009–2011 Outcome 2011–2013	Starters 2009–2011 Outcome 2012–2014	Starters 2009–2010 Outcome 2013–2014
DI benefit (yes) in %	Kernel, levels	–0.1 (–0.28)	–2.6*** (–2.94)	–4.0*** (–3.93)	–6.1*** (–4.13)
	Kernel, DiD	–1.0 (–1.29)	–3.4*** (–3.45)	–4.9*** (–4.21)	–7.8*** (–4.92)
	Radius, DiD	–0.2 (–0.17)	–2.4* (–1.69)	–3.1** (–1.97)	–6.9*** (–3.26)
Monthly main DI benefit in CHF	Kernel, levels	–42*** (–3.23)	–85*** (–5.2)	–108*** (–5.78)	–142*** (–6.10)
	Kernel, DiD	–51*** (–3.92)	–94*** (–5.67)	–118*** (–6.24)	–146*** (–6.26)
	Radius, DiD	–51** (–2.74)	–93*** (–4.17)	–114*** (–4.59)	–121*** (–3.77)
DI benefit entitlement in %	Kernel, levels	–2.3*** (–3.11)	–4.7*** (–5.05)	–6.1*** (–5.68)	–7.9*** (–5.97)
	Kernel, DiD	–2.9*** (–3.78)	–5.3*** (–5.52)	–6.7*** (–6.14)	–8.3*** (–6.05)
	Radius, DiD	–2.7** (–2.53)	–5.1*** (–4.06)	–6.2*** (–4.42)	–6.9*** (–3.81)
Monthly total DI benefit in CHF (main + child's benefit)	Kernel, levels	–64*** (–3.97)	–117*** (–5.66)	–143*** (–6.02)	–183*** (–6.27)
	Kernel, DiD	–63*** (–3.88)	–115*** (–5.68)	142*** (–6.12)	–176*** (–6.06)
	Radius, DiD	–61** (–2.55)	–110*** (–3.90)	–129*** (–4.05)	–119*** (–2.86)
Monthly SB per case in CHF	Kernel, levels	–65*** (–3.12)	–73*** (–3.01)	–91*** (–3.61)	–117*** (–3.86)
	Kernel, DiD	–71*** (–3.55)	–79*** (–3.34)	–97*** (–3.95)	–125*** (–4.08)
	Radius, DiD	–57** (–2.19)	–59* (–1.97)	–85** (–2.58)	–81** (–2.15)
Number of participants	908	908	908	578	
Number of controls	40,710	40,710	40,710	27,140	

p* < 0.10, *p* < 0.05, ****p* < 0.01. See Table 4 for definitions of the outcome variables

13% points higher than in the control group. As a result, the average annual income from paid employment in *t* + 2 is CHF 2750 higher, corresponding to a relative increase of approximately 37%.⁸ In *t* + 3 the ATT effect on income is CHF 3975 (approx. +56%). The amount of the monthly main DI benefit is reduced by 3.8% in *t* + 1 compared to the control group and by 11% in *t* + 4, i.e. by a monthly

amount of CHF 51 in *t* + 1 and CHF 146 in *t* + 4. This equates to annualized amounts of CHF 612 and CHF 1752, respectively. For the monthly total DI benefit, the decreases are CHF 63 in *t* + 1 and CHF 176 in *t* + 4, which leads to annualized amounts of CHF 756 and CHF 2112.

At least until *t* + 4 it can be seen that the favorable ATT effects are not only temporary. For example, the ATT effect on the DI benefit entitlement in % increases from –2.9% points in *t* + 1 to –8.3% points in *t* + 4. The monthly supplementary benefits are reduced by CHF 71 (= 12%) in *t* + 1 up to CHF 125 (= 19%) in *t* + 4.

The Web Appendix includes sensitivity analyses with regard to (1) unobserved confounders (selection on

⁸ This relative increase of 37% can be roughly calculated as follows: from the graph in the bottom-left of Fig. 3 it can be seen that the counterfactual income is approximately CHF 7500. CHF 2750/CHF 7500 is approximately 37%

Table 7 Estimated ATT-effects on income—*t* stats in parentheses

Outcome variable	Matching method	<i>t</i> + 1	<i>t</i> + 2	<i>t</i> + 3
		Starters 2009–2011 Outcome 2010–2012	Starters 2009–2011 Outcome 2011–2013	Starters 2009–2010 Outcome 2012–2013
Income earned from paid employment (yes) in %	Kernel, levels	8.4*** (5.42)	12.1*** (7.33)	11.7*** (6.19)
	Kernel, DiD	9.3*** (5.27)	13.0*** (5.84)	12.2*** (6.50)
	Radius, DiD	9.8*** (4.81)	12.0*** (4.62)	13.2*** (4.83)
Annual income earned from paid employment in CHF	Kernel, levels	1549** (2.83)	2313*** (4.16)	3479*** (4.85)
	Kernel, DiD	1983*** (3.85)	2747*** (4.95)	3975*** (5.37)
	Radius, DiD	2198*** (3.64)	2922*** (4.57)	4017*** (4.75)
Number of participants		908	908	578
Number of controls		40,710	40,710	27,140

p* < 0.10, *p* < 0.05, ****p* < 0.01. See Table 4 for definitions of the outcome variables

unobservables/omitted variables), (2) the bandwidth choice in case of the kernel-based matching estimator, and (3) approaches to impose common support. All in all, the results seem to be quite robust.

Assessment of Costs and Benefits

Based on the estimated ATT effects, this section will present a simple cost–benefit analysis from the Swiss social security system’s perspective. Due to the inherent uncertainty an analysis like this can be only a rough guideline. The question is whether the pilot project was an advantageous investment for the social security system. This will be the case, if the initial expenditures of around CHF 8819 (=EUR 8038 or USD 8970)⁹ per participant are overcompensated by a future (discounted) reduction in DI benefits. Due to the limited scale and temporary nature of the measure, general equilibrium and macroeconomic effects can be neglected in the analysis.

Any assessment of future reductions in payments must be based on arbitrary assumptions since the empirical estimates of the ATT effects are only available for the first 4 years after the individual start of the program. Moreover, estimated ATT effects for *t* + 4 are based on a reduced number of participants (578 instead of 908). However, the entire period until the start of old-age pension is of relevance in this context. Given the participants’ average age of 45, that

amounts to a 20-year period. It is necessary to apply two assumptions already mentioned in the section “[Data and Sample Selection](#)”: (1) the treatment does not affect mortality; (2) the treatment does not affect the retirement age. While the former assumption seems reasonable, the second assumption is problematic. However, since it is not possible to estimate the effect of the treatment on the retirement age there is no alternative to this assumption.

Various scenarios are applied in order to determine what effects the treatment would have on total DI benefits over the period until the participants reach pensionable age. The scenarios can be distinguished on the basis of whether they examine the permanent (lasting) ATT effects or only temporary ATT effects (over some years). Figure 4 provides a graphical representation of these scenarios, which are explained in greater detail in the Section IV in the Web Appendix .

Furthermore, since the ATT effects on total taxable income are estimated, it is possible to calculate the additional social security contributions resulting from higher income. The total contribution rate, which is assumed to be constant in the future, is 12.5%.¹⁰ However, due to issues involving data availability, as described in the section “[Data and Sample Selection](#)”, the ATT effects are only estimated up to *t* + 3. Hence, the scenarios described above have to be adjusted (see the Web Appendix). While it is possible to

⁹ Exchange rates at 16-March-2016.

¹⁰ This number is the sum of the contribution rates of the AHV (8.4%), IV (1.4%), EO (0.5%), and the ALV (2.2%). Source: Federal Social Insurance Office (FISO), Switzerland

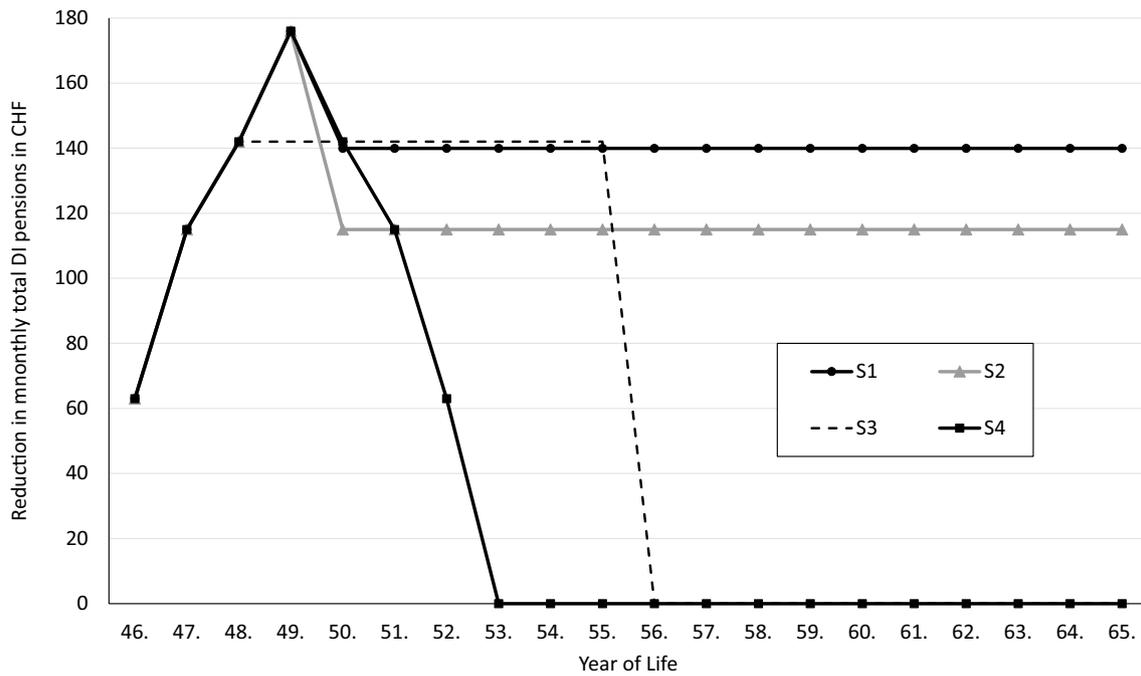


Fig. 4 Graphical representation of the scenarios assumed for the cost–benefit analysis. $t + 1$ corresponds to the 46th year of life. The first four years (46, 47, 48, 49) correspond to the kernel-based DiD matching estimates of the ATT effect in Table 6

estimate the effects on social security contribution, this is not possible regarding income tax (or other taxes), because not all necessary information (such as further incomes) is available.

Working with price-adjusted (real) outcome variables implies that real interest rates have to be used for discounting future payments in the cost–benefit analysis. Choosing a (real) discount rate is the next arbitrary assumption. Historical data on the bank lending rate minus inflation measured by the GDP deflator provided by the World Bank [45] indicate for Switzerland an average of 2.8% for the period 1981–2014. However, this may not be the appropriate discount rate for government programs. At present (March 2016), the (nominal) returns on Swiss government bonds are extremely low and even negative for some maturities. Given the low returns on Swiss bonds and the discussion of “secular stagnation” [46], 3.0% seems to be a reasonable upper bound for the average real interest rate for a 20-year time period commencing from the program. For the four scenarios introduced above and for three different real interest rates ($r = 1.0, 2.0, 3.0\%$) the current values of the expected future reductions in DI benefits and supplementary benefits as well as the current value of expected future additional social security contributions are calculated.

Table 8 shows the results, which are rounded to CHF 100 in order to avoid spurious accuracy. Comparing the social security system’s investment of CHF 8819 per participant with the current values of expected reductions in

Table 8 Current values of expected reductions in total DI benefit payments and additional social security contributions by scenario and discount rate (per participant in CHF)

Scenario	Real discount rate		
	1.0%	2.0%	3.0%
S1			
Decrease in total DI benefits	29,500	26,700	24,200
Decrease in supplementary benefits	20,800	18,800	17,100
Increase in social security contributions	6800	6200	5600
Sum	57,100	51,700	46,900
S2			
Decrease in total DI benefits	17,000	15,500	14,200
Decrease in supplementary benefits	16,600	15,100	13,700
Increase in social security contributions	4700	4300	3900
Sum	38,300	34,900	31,800
S3			
Decrease in total DI benefits	14,900	14,100	13,300
Decrease in supplementary benefits	10,500	9900	9400
Increase in social security contributions	3400	3200	3100
Sum	28,800	27,200	25,800
S4			
Decrease in total DI benefits	9400	9100	8700
Decrease in supplementary benefits	7100	6900	6600
Increase in social security contributions	1600	1500	1500
Sum	18,100	17,500	16,800

DI benefits and supplementary benefits, it becomes obvious that the pilot project led to a net benefit for the social security system. This is even more clearly the case, when adding together all three positions (DI benefit, supplementary benefits, social security contributions). Based on this, the expected net benefit of the social security system is between CHF 8000 and CHF 48,300 per participant.

Summary and Conclusions

During 2009–2013 a pilot project was carried out in Zurich which aimed to increase the income of DI benefit recipients in order to reduce their entitlement to DI benefits. The project consisted of placement coaching carried out by a private company that specialized in this field. It was exceptional with respect to three aspects: (1) it did not include any formal training and/or medical aid; (2) the coaches did not have the possibility of providing additional financial incentives or making use of “threat effects”; and (3) due to performance bonuses, the company not only had incentives to bring the participants in (higher paid) work, but also to keep them there for 52 weeks.

This paper estimates the medium-run effects of the pilot project and assesses the net benefit for the Swiss social security system. For this purpose, an administrative panel data set is analyzed. Three possibly important groups of variables are missing: information on educational background, personality traits, and further income sources. All three may affect the outcome as well as selection into the treatment. In order to address the possibly resulting problem of confounders/selection on unobservables a DiD approach is applied in addition to matching. Moreover, the sensitivity of the results with regard to this problem is simulated in Section III.1. in the Web Appendix. For the balancing of observable variables kernel-based matching on the propensity score as well as radius matching with bias adjustment and additional variables in the Mahalanobis distance are applied. In terms of “balancing tests” kernel-based matching outperforms the radius matching estimator. The estimated treatment effects do not differ significantly with the method.

All the estimation results indicate that, on average, the project was effective for the participants in the sense that the DI benefit amounts paid to them could be lowered by about 10% by raising their levels of paid employment and income. Up to 4 years after the starting year there is no indication that this positive effect is only temporary. Based on different scenarios and different discount rates, the current value of future reductions in benefit payments and additional social security contributions is calculated. The expected mean long-run benefits exceed the mean costs by 1.9–6.5 times. Subtracting the costs of the project, the expected net benefit

of the social security system is between CHF 8000 (=EUR 7290 or USD 8136) and CHF 48,300 (=EUR 44,016 or USD 49,121) per participant.¹¹

How can these favorable results be explained? First of all, one should note, that in general job search assistance programs are relatively effective according to a lot of empirical evidence. One explanation for this effectiveness and for the results of this paper is the absence of lock-in effects, which often hampers the effectiveness of training programs [15]. This is especially true for the coaching program evaluated here. Figure 3 in the section “Pre-program Outcome Variables” indicates that the counterfactual outcomes are almost constant after the start of the treatment. Hence, if the participants did not participate, they would not be able to improve their employment situation, although they could participate in “traditional” VR measures provided by the public DI office. Moreover, the findings of this paper are in line with previous studies on active labor market policies for disabled workers. The majority of evaluations show positive employment and earnings effects. Markussen and Røed [14] found that a strategy focusing on rapid placement in the regular labor market is superior.

An interesting policy implication is, that under favorable labor market conditions it is possible to enhance the employment prospects of disabled persons with a relatively inexpensive intervention that does not include any explicit investments in human capital (vocational training), health support, financial incentives or threat effects. It is difficult to derive further policy recommendations from these results. Quantitative evaluation studies are able to answer the question whether programs are effective or not. However, they are usually not able to identify the reasons for the effectiveness of the analyzed programs. One possible important reason for the effectiveness stressed by the private company, which carried out the pilot project, is the importance of developing self-esteem and motivation of participants. A second reason may be the incentives for the company due to performance bonuses. It not only had incentives to bring the participants into (higher paid) work, but also to keep them there for 52 weeks. Finally, it may be easier for coaches outside the public DI system to establish a relationship of trust with the recipients since they do not decide on entitlements to benefits.

Acknowledgements The author would like to thank the anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper.

Funding This paper is based on the project “Evaluation Pilotprojekt Ingeus—berufliche Wiedereingliederung von Rentenbeziehenden der Invalidenversicherung” funded by the Federal Social Insurance Office

¹¹ Exchange rates at 16-March-2016.

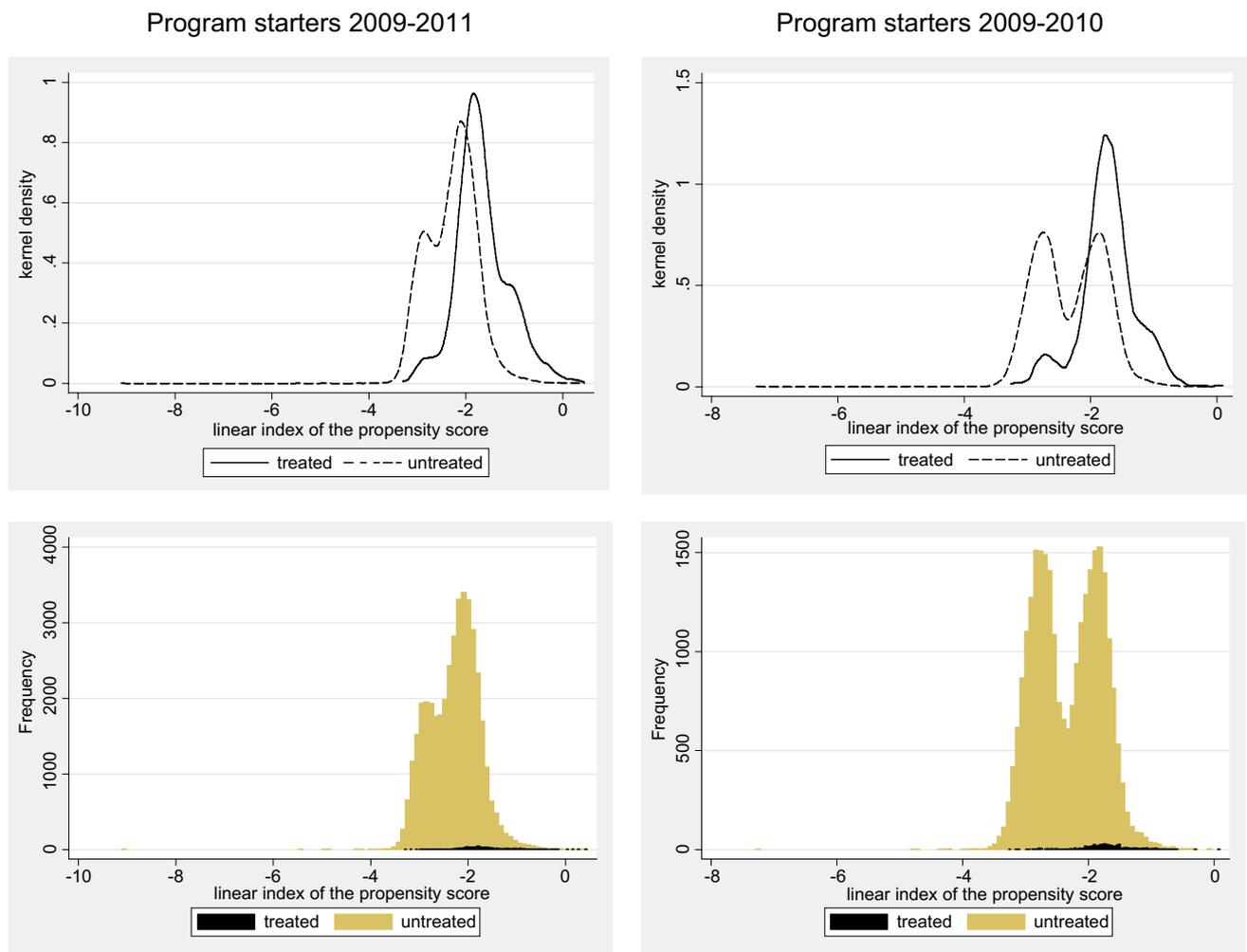


Fig. 5 Kernel density estimate and histogram (frequency) of the propensity score estimates

(FISO), Switzerland. It does not necessarily reflect the opinions and views held by the FISO.

Compliance with Ethical Standards

Conflict of interest Author Tobias Hagen declares that he has no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Appendix

See Fig. 5.

References

1. Burkhauser RV, Daly MC, McVicar D, Wilkins R. Disability benefit growth and disability reform in the US: lessons from other OECD nations. *IZA J Labor Policy*. 2014;3(1):1–30.
2. Organisation for Economic Co-operation and Development. *Sickness, disability and work: breaking the barriers, a synthesis of findings across OECD countries*. Paris: OECD Publishing; 2010.
3. Bound J, Burkhauser RV. Economic analysis of transfer programs targeted on people with disabilities. *Handb Labor Econ*. 1999;3(3):3417–3528.
4. Wittenburg D, Mann DR, Thompkins A. The disability system and programs to promote employment for people with disabilities. *IZA J Labor Policy*. 2013;2(1):4.
5. Maestas N, Mullen KJ, Strand A. Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt. *Am Econ Rev*. 2013;103(5):1797–1829.

6. Büttler M, Deuchert E, Lechner M, Staubli S, Thiemann P. Financial work incentives for disability benefit recipients: lessons from a randomised field experiment. *IZA J Labor Policy*. 2015;4(1):1–18.
7. Delin BS, Hartman EC, Sell CW. Given time it worked: positive outcomes from a ssdi benefit offset pilot after the initial evaluation period. *J Disabil Policy Stud*. 2015;26(1):54–64.
8. Weathers RR, Hemmeter J. The impact of changing financial work incentives on the earnings of Social Security Disability Insurance (SSDI) beneficiaries. *J Policy Anal Manag*. 2011;30(4):708–728.
9. Frölich M, Heshmati A, Lechner M. A microeconomic evaluation of rehabilitation of long-term sickness in Sweden. *J Appl Econ*. 2004;19(3):375–396.
10. Heshmati A, Engström LG. Estimating the effects of vocational rehabilitation programs in Sweden. In: Lechner M, Pfeiffer F, editors. *Econometric evaluation of labour market policies*. ZEW economics studies 13. New York: Physica-Verlag; 2001. pp 183–210.
11. Campolieti M, Gunderson MK, Smith JA. The effect of vocational rehabilitation on the employment outcomes of disability insurance beneficiaries: new evidence from Canada. *IZA J Labor Policy*. 2014;3(1):1–29.
12. Aakvik A, Heckman JJ, Vytlačil EJ. Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs. *J Econ*. 2005;125(1):15–51.
13. Dean D, Pepper J, Schmidt R, Stern S. The effects of vocational rehabilitation for people with cognitive impairments. *Int Econ Rev*. 2015;56(2):399–426.
14. Markussen S, Røed K. The impacts of vocational rehabilitation. *Labour Econ*. 2014;31(1):1–13.
15. Brown AJ, Koettl J. Active labor market programs—employment gain or fiscal drain? *IZA J Labor Econ*. 2015;4(1):1–36.
16. Card D, Kluve J, Weber A. Active labour market policy evaluations: a meta-analysis. *Econ J*. 2010;120(548):F452–F477.
17. Thomsen S. Job search assistance programs in Europe: evaluation methods and recent empirical findings. FEMM Working Paper No. 18. Magdeburg: Otto-von-Guericke University; 2009.
18. Wunsch C. How to minimize lock-in effects of programs for unemployed workers. *IZA World Labor*. 2016. <https://doi.org/10.15185/izawol.288>.
19. Høgelund J, Holm A. Case management interviews and the return to work of disabled employees. *J Health Econ*. 2006;25(3):500–519.
20. LaLonde RJ. Evaluating the econometric evaluations of training programs with experimental data. *Am Econ Rev*. 1986;76(4):604–620.
21. Dehejia RH, Wahba S. Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. *J Am Stat Assoc*. 1999;94(448):1053–1062.
22. Smith J, Todd P. Does matching overcome LaLonde’s critique of nonexperimental estimators? *J Econ*. 2005;125(1–2):305–353.
23. Fredriksson P, Johansson P. Dynamic treatment assignment: the consequences for evaluations using observational data. *J Bus Econ Stat*. 2008;26(4):435–445.
24. Sianesi B. An evaluation of the Swedish system of active labor market programs in the 1990. *Rev Econ Stat*. 2004;86(1):133–155.
25. Sianesi B. Differential effects of active labour market programs for the unemployed. *Labour Econ*. 2008;15(3):370–399.
26. Biewen M, Fitzenberger B, Osikominu A, Waller M. The effectiveness of public sponsored training revisited: the importance of data and methodological choices. *J Labor Econ*. 2014;32(4):837–897.
27. Leuven E, Sianesi B. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. 2003. This version 4.0.11. <http://ideas.repec.org/c/boc/bocode/s432001.html>.
28. Imbens GW. Matching methods in practice: three examples. *J Hum Resour*. 2015;50(2):373–419.
29. Galdo JC, Smith J, Black D. Bandwidth selection and the estimation of treatment effects with unbalanced data. *Ann Econ Stat*. 2008;91–92:189–216.
30. Silverman BW. *Density estimation for statistics and data analysis*. London: Chapman & Hall; 1986.
31. Lechner M, Miquel R, Wunsch C. Long-run effects of public sector sponsored training in West Germany. *J Eur Econ Assoc*. 2011;9(4):742–784.
32. Huber M, Lechner M, Steinmayr A. Radius matching on the propensity score with bias adjustment: tuning parameters and finite sample behavior. *Empir Econ*. 2015;49(1):1–31.
33. Caliendo M, Mahlstedt R, Mitnik O. Unobservable, but unimportant? The influence of personality traits (and other usually unobserved variables) for the evaluation of labor market policies. *IZA Discussion Paper No. 8337*; 2014.
34. Lechner M, Wunsch C. Sensitivity of matching-based program evaluations to the availability of control variables. *Labour Econ*. 2013;21(1):111–121.
35. Heckman JJ, Smith JA. The pre-programme earnings dip and the determinants of participation in a social programme. Implications for simple programme evaluation strategies. *Econ J*. 1999;109(457):313–348.
36. Heckman JJ, LaLonde RJ, Smith JA. The economics and econometrics of active labor market programs. *Handb Labor Econ*. 1999;3(1):1865–2097.
37. Dehejia R. Practical propensity score matching: a reply to Smith and Todd. *J Econ*. 2005;125(1):355–364.
38. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar Behav Res*. 2011;46(3):399–424.
39. Millimet DL, Tchernis R. On the specification of propensity scores, with applications to the analysis of trade policies. *J Bus Econ Stat*. 2009;27(3):397–415.
40. Ashenfelter O. Estimating the effect of training programs on earnings. *Rev Econ Stat*. 1978;60(1):47–57.
41. Lechner M. A Note on the common support problem in applied evaluation studies. *Ann Econ Stat*. 2008;91–92:217–235.
42. Lee WS. Propensity score matching and variations on the balancing test. *Empir Econ*. 2013;44(1):47–80.
43. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat*. 1985;39(1):33–38.
44. Rubin DB. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Serv Outcomes Res Methodol*. 2001;2(3–4):169–188.
45. World Bank. *World development indicators*. 2015. <https://data.worldbank.org>. Accessed 09 Feb 2016.
46. Baldwin R, Teulings C. *Secular stagnation: facts, causes and cures*. London: Centre for Economic Policy Research-CEPR; 2014.