FRANKFURT UNIVERSITY OF APPLIED SCIENCES

Working Paper Series: Business and Law

Working Paper No. 10

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

Tobias Hagen

März 2016

Fachbereich 3
Wirtschaft und Recht | Business and Law

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

TOBIAS HAGEN*

This version: March 24, 2016

Abstract

This paper evaluates a placement coaching program implemented in Zurich during 2009–2013 that focused on the reemployment of persons drawing disability insurance (DI) benefits. A private company was commissioned to implement the program. Kernel-based matching and radius matching with bias adjustment estimators combined with difference-in-differences are applied to administrative panel data. The estimates point to a successful project in terms of a reduction in DI benefits and an increase in income even in the medium-run. A simple cost–benefit analysis suggests that the project was a profitable investment for the social security system. Sensitivity analyses indicate that the results are robust to confounders and further specification issues. An interesting policy implication is that 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.

* Frankfurt University of Applied Sciences, Nibelungenplatz 1, 60318 Frankfurt am Main, Germany (e-mail: tobiashagen@email.de).

I. 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 (Burkhauser et al. 2014, Organisation for Economic Co-operation and Development (OECD) 2010, Bound and Burkhauser 1999). 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 (Burkhauser et al. 2014). As explained by Wittenburg et al. (2013) employment programs for disabled workers must help them to overcome substantial employment barriers: (1.) the loss of human capital due to the disability and the prolonged separation from the workforce; (2.) 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 earnings increases (see, e.g., Maestas et al. 2013). The latter aspect is of particular importance for Switzerland (see Bütler et al. 2015).

The traditional active labor market policy for this target group is vocational rehabilitation (VR), often associated with human capital investments 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 to give them the qualifications need for new jobs (OECD 2010).

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. In this paper a microeconometric evaluation as well as a simple 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 for recipients of DI benefits is scarce. Firstly, there are several studies of the effectiveness of VR measures 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 for recipients of DI benefits.

Starting with evaluation studies on VR measures, using Propensity Score Matching Frölich et al. (2004) do not find positive effects of participation in such programs compared to non-participation in Sweden. However, their results show that workplace training is superior to the other rehabilitation measures. In contrast to this result, based on a bivariate probit model Heshmati and Engström (2001) find that participation in VR programs in Sweden has positive effects on the health status as well as the rate of return to work. Recently, using several matching and weighting estimators, Campolieti et al. (2014) find that a VR program in Canada improves the labor market outcomes of women, but not men. This study also attempt 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) exceed the costs. Aakvik et al. (2005) evaluate VR programs for female applicants in Norway. They find a negative effect of the training program on employment prospects. The estimated effects are larger 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 are infrequently selected into the measure. For a subgroup of workers with cognitive impairments in the state of Virginia, Dean et al. (2015) are able to evaluate the effects of multiple services of VR using an instrumental variable approach.[1] They find large positive long-run (three to nine years) effects on employment and earnings. Both the short- and long-run mean labor market effects are estimated to be positive for diagnosis and evaluation, training, education and other services, but negative for restoration and maintenance. The mean long-run benefits exceed the mean costs by four to six times. For Norway, Markussen and Røed (2014) evaluate four different VR programs for temporary DI claimants. Based on longitudinal administrative data they use local variations in the policy strategies to estimate the impact of these strategies on the participants' future employment and earnings performance. Overall they estimate positive effects. However, they find that a strategy focusing on rapid placement in the regular labor market is 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,

---

[1] An explanation of the different services is given in Dean et al. (2015) on page 403.

it can be said that these kinds of measures seem to be 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 (2015) these measures may work not only 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 a part of the program. This should be kept in mind when looking at the previous empirical evidence. Using a meta-analysis, Card et al. (2010) find that job search assistance programs yield relatively positive effects compared to other measures. Thomsen (2009) reviews studies related to nine European countries and finds that job search assistance programs decrease unemployment duration and increase employment rates. This is confirmed by a review provided by Brown and Koettl (2015). They conclude that "measures improving labor market matching" are cost-effective and may have significant short-run effects. They also conclude that lock-in effects may be minimal, which seems to be important for the understanding of this paper's results. The programs should be targeted at persons with bad employment prospects at the beginning of their unemployment spell and at long-term unemployed persons. Moreover, they find that these programs are most effective during recoveries.

Finally, there is only one empirical study evaluating a program with some similarities to the pilot project in Zurich. Høgelund and Holm (2006) evaluate the effect of case management interviews (CMIs) performed by social caseworkers on the probability of a return to work by disabled employees. Based on instrumental variables and a competing hazard rate model, they find that CMIs increase the probability of returning to work for the pre-sick leave employer, but has 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 all of the VR instruments available that may help the DI benefit recipients. This is definitely not the case in the evaluated pilot project in this paper.

To the best of my knowledge, this paper presents the first econometric evaluation of a placement coaching program for DI beneficiaries. Placement coaching programs may be of special interest for policymakers, as they are relatively inexpensive in comparison to formal training programs (see Brown and Koettl, 2015). 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 (four years after the program start). The cost–benefit analysis indicate 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 to 6.5 times.

The remainder of the paper is organized as follows: the Swiss DI system is briefly outlined in Section II. Section III 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 Section IV. Section V 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 Section VI. In Section VII a sensitivity analysis provides information on the robustness of the estimated effects. Section VIII 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.

II. 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.[2] Disability is defined as a decline in the ability to earn a living or in the ability to accomplish of daily tasks, such

---

[2] This section is based on http://www.zas.admin.ch/org/00858/00861/index.html?lang=en.

as housework, resulting from physical, psychological or mental health problems. To qualify as a disability this incapacity must last at least one year. When judging if an inability to earn a living is present it does not matter what the causes of the health depreciation are. Furthermore, there is only an inability to earn a living if one is not able to overcome it for "objective reasons".

Insured persons who have paid contributions for at least three 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 *disability degree*) and after a waiting period of one 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 (see Bütler et al., 2015). 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.

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 obs.	40,710	908

Notes: The descriptive statistics are based on the estimation sample described in Section IV.

III. The Pilot Project and the Selection Mechanism

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 advisers 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 advisers used publicly available job advertisements. The company did not have its own vacancy database. However, due to the relatively favorable labor market conditions in Zurich, this may be a minor issue. The advisers did not have the possibility of providing any financial incentives, or to sanction their clients for a lack of effort.

The "placement phase" lasted for a maximum of 12 months. Those who dropped out prematurely were given the option of starting the measure again. "Dropout" here means that participants did not show up at appointments with their advisers or that they indicated that they did not want to participate anymore. 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 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 three months). Note that this is not a special feature of the project, but rather is the regulated process for all DI benefit recipients.

In addition to the sustainability bonuses of CHF 3,000 paid for every participant placed in a new job for a period of 26 weeks (or double that amount for 52 weeks), the amount invested in the pilot project by the DI scheme comprised a lump sum per case of CHF 6,000 and overall set-up costs of CHF 2.28 million. The total cost per participant was CHF 8,819.

In summary, the pilot project was exceptional with respect to three aspects: (1.) it did not include any formal training and/or medical aid; (2.) the coaches/advisors did not have the possibility of

providing additional financial incentives or making use of "threat effects"; and (3.) 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.
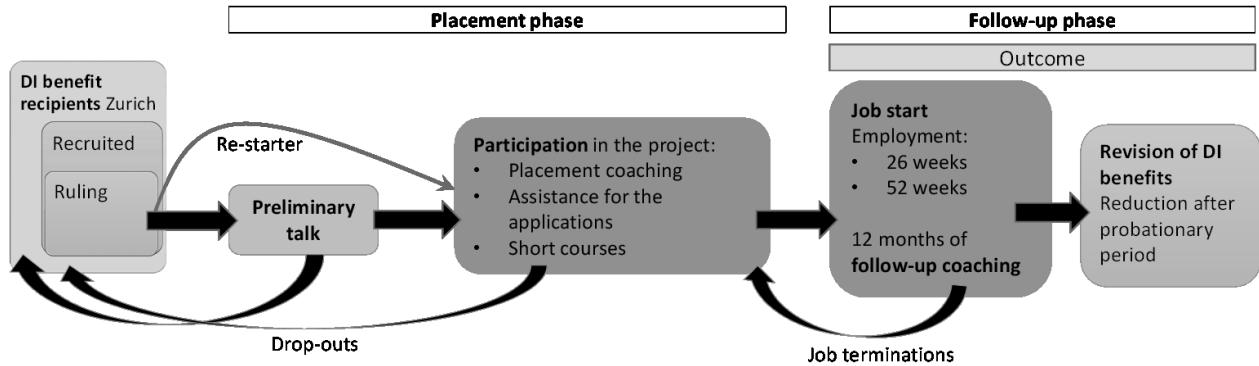
FIGURE 1—FLOWCHART OF THE PILOT PROJECT

Before DI benefit recipients could take part in the pilot project, they had to complete the three-stage process depicted in the far-left of Figure 1 and described in the following paragraph. 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. This means, that 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. 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*, 1,368 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.

IV. Data and Sample Selection

The whole analysis is based on administrative data that the Federal Social Insurance Office had gathered from six data registers. For the period 2000–2014 these data include for all DI benefit recipients in the Canton of Zurich 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 for December 2009. The total number of recipients at the end of the year 2009 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.

TABLE 2 —NUMBER OF INDIVIDUALS: TREATED AND POTENTIAL CONTROLS IN DEC. 2009

All DI benefit recipients in Zurich	49,951
Persons recruited*	14,878
Persons who received rulings*	1,037
Participants (2009-2011)*	908

Notes: * Individuals who died until December 2014 and people who transitioned to the old-age pension system are excluded.

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.

TABLE 3—NUMBER OF PROGRAM-STARTERS AND NON-PARTICIPANTS BY YEAR

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

As explained in greater detail below, the applied approach strives for annual accuracy only. The potential control group is based on 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.

TABLE 4—OUTCOME VARIABLES

Outcome variable	Period	Adjusted for	Notes
Monthly main DI benefit in CHF	Dec. 2000 – Dec. 2014	DI benefit increases, every second year)	Measured in Dec. each year
Monthly total DI benefit in CHF (main benefit and child's benefit)	Dec. 2000 – Dec. 2014	DI benefit increases, every second year	Measured in Dec. each year
DI benefit entitlement in %	Dec. 2000 – Dec. 2014		measured in Dec. each year; see Table 1 for an explanation
Monthly supplementary benefits (SB) per case in CHF	Dec. 2000 – Dec. 2014	Consumer price index, annual	Total payment for households
DI benefit recipient (yes) in %	Dec. 2000 – Dec. 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

The outcome variables are shown in Table 4. The data on DI benefit entitlements and payments are measured in December each year for the period 2000–2014, and therefore the whole analysis can only be based on an annual frequency. The outcome variables measured in CHF are adjusted either for DI benefit increases or consumer price inflation to the base year 2009. Price-adjusted outcome variables can be directly compared over time. For example, a single person without any change in her DI benefit entitlement has a constant price-adjusted benefit over time.

The following conditioning variables are available for the analysis (see Table 6 in Section VI.B.1 for sample means):

- Socio-demographic characteristics: Standard variables such as *age*, *gender*, *nationality*, *civil status* and *number of children* are available. The latter variable is not measured directly, but is based on the *number of child DI benefits*. These are paid for children under the age of 18, or until the completion of education, with a maximum age of 24. 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 the administrative data. 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 (*impairment of general condition*, *behavioral disorders*, *multiple mental disorders*, *impairment of the trunc*, *multiple mental and physical disorders*) 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 since the year 2000. Furthermore, previous *participation in VR* measures under the DI system can be identified. The *daily allowance* of the DI system also indicates that a person participated in a VR measure. The receipt of a *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*. This information is available in the data from the year 2000 onwards.

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 in the estimated effects. 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 (1986), Dehejia and Wahba (1999, 2002), Smith and Todd (2005) and others.

V. Econometric Approach

A. Basics

The goal is to estimate the 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 participation in the measure, irrespective of the "completion" (i.e. including dropouts from the program). The counterfactual is non-participation in the program. The non-participants are defined here as individuals who were selected by the DI office as potential participants (were recruited), but decided not to participate (see Figure 1 and 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 (2008) as well as Sianesi (2004, 2008) 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 (Biewen et al. 2014). 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 is 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, in my view the standard static approach can be safely applied. That means, that non-participants can be defined as individuals who never participate in that program, but do possibly participate in another traditional VR measure.

In concrete terms, the econometric approach used to estimate the average treatment on the treated (ATT) effect is propensity score matching. Here "psmatch2" implemented in STATA by Leuven and Sianesi (2003) is applied. The estimation procedure is as follows: the propensity score equation for the participation (C) 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.[3] Note that 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 $\hat{e}_1(X)$ und $j = 1, \dots, N_0$ and untreated individuals $\hat{e}_0(X)$ is predicted. The common support condition is fulfilled, if for every treated i the predicted propensity score is "sufficiently" overlapped by the predicted propensity scores of untreated individuals.

After matching, the ATT effect is calculated as $ATT = \frac{1}{N_1} \sum_{i=1}^{N_1} \left(Y_{1i} - \sum_{j=1}^{N_0} w(i,j) Y_{0j} \right)$, 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. Since it leads to the best results

---

[3] As discussed in greater detail in Section V.B, N_0=40,710 can be explained by the fact that the 13,570 non-participants can be used for every year 2009, 2010 and 2011 as controls.

in terms of balancing pre-treatment differences in the outcome variables and the conditioning variables between the treatment and the control group, kernel-based matching using the Epanechnikov kernel is applied here. This kernel has the advantage of attaching a weight of zero to control observations outside the bandwidth (see Galdo et al. 2008).

In recent years there has been an increasing insight into optimal bandwidth choice in kernel matching (see, e.g., the discussion in Galdo et al. 2008 and Biewen et al. 2014). Silverman's (1986) rule of thumb suggests a bandwidth value of 0.06, which will be used here.[4] Surprisingly, and in contrast to the literature, the estimated ATT effects here are relatively robust to changes in the bandwidth (see Section VII.B).

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. (2011) propose a distance-weighted radius matching with bias adjustment. Huber et al. (2015) implement this estimator in STATA with the command "radiusmatch". 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. (2015) find that including the most important covariates (on top of the propensity score) in the matching algorithm (via the Mahalanobis metric) leads to better results in terms of decreasing the selection bias. Here the variables *gender*, *year of the treatment* (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. (2015), the default values of the STATA command 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, disregarding the fact the propensity score is estimated.

In order to eliminate the possible bias due to selection on unobservables, the matching procedure is extended by difference-in-differences (DiD) (Heckman et al. 1998b). With DiD the ATT effect is estimated as $ATT = \frac{1}{N_1}\sum_{i=1}^{N_1}\left(\left(Y_{1i\tau} - Y_{1i,t-1}\right) - \sum_{j=1}^{N_0} w(i,j)\left(Y_{0j\tau} - Y_{0j,t-1}\right)\right)$, 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

---

[4] This method of bandwidth choice is applied inter alia by Heckman et al. (1998b).

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. For the example of the ATT effect in $t + 1$, this means for the treatment group $(Y_{1,t+1} - Y_{1,t-1})$ and for the control group $(Y_{0,t+1} - Y_{0,t-1})$. Smith and Todd (2015) find 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. (2014) find 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 Lechner and Wunsch 2013). 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.

However, DiD and conditioning on pre-treatment outcome varwiables 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 (Heckman and Smith 1999, Heckman et al. 1999). 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. Sensitivity analyses on the potential bias due to selection on unobservables / confounders are presented in Section VII.A.

B. 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 (Imbens 2015), 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. The literature also suggests 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 (Dehejia and Wahba 1999, Dehejia 2005). Moreover, as panel data are available it is possible to include lagged values of time-varying X (including Y) up to $t - 9$. 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 (see Caliendo and Kopeinig 2008, Millimet and Tchernis 2009, Austin 2011, Imbens 2015).

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" (Heckman and Hotz 1989) is shown in Section VI.B.1 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 Section V.A, 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 arise from the participation. Also, the empirical data in Section VI.B.1 do not show a "dip" in $t-1$ or $t-2$ (Ashenfelter 1978; Heckman and Smith 1999). 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. There are three possible approaches, as set out in the following paragraphs.

 (1.) *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 were 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, the starting year, 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 controls to 13,570 per year (Table 3).

(2) *Ignore t and allow that all 40,710 non-participants are potential controls*. This would be a valid approach if changes in the labor market situation over time were not relevant.

(3) *Define the year 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 pre-treatment outcome variables. However, it comes with the cost that t is not identical for all treated and corresponding controls.

---

[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 www.amstat.ch.

Without changing anything explained above, matching is here not on the propensity score, but on the underlying linear index (see, e.g., Lechner 2008). 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. The estimation of ATT effects by kernel-based and radius matching and the estimation of the corresponding standard errors was described in the previous subsection.

VI. Empirical Results

A. Propensity Score

Table 5 shows the estimation results of the propensity score probit for both samples. Due to high collinearity the coefficients have no causal interpretation (Imbens 2015). Nevertheless, there is one interesting results, which should be mentioned. The individual recent labor market history seems particularly important since variables such as *daily allowance of the unemployment insurance* and *participation in VR measures* have statistically significant effects. However, they may simply serve as proxies for the unobserved variables.

TABLE 5 — ESTIMATES OF PROPENSITY SCORE, PROBIT MODEL

	Coefficient (t-stat.)	
Conditioning variables Lagged values (previous calendar year), unless specified differently.	Program starters 2009-2011	Program starters 2009-2010
Years receiving DI benefit (since 2000)	-0.0278*** (-4.36)	-0.0131 (-1.58)
Age	0.0497*** (3.60)	0.0743*** (4.19)
Age squared / 1,000	-0.736*** (-4.57)	-0.0010*** (-4.93)
Woman# (ref.: man)	-0.0047 (-0.15)	-0.0241 (-0.63)
Type of diseases (ref.: congenital defects)		
mental#	-0.101 (-1.17)	-0.0645 (-0.53)
nervous system #	-0.140 (-1.37)	0.0303 (0.36)
musculo-skeletal #	-0.0966 (-1.10)	0.00238 (0.02)
other#	-0.196** (-2.04)	0.0501 (0.56)
injuries#	-0.127 (-1.32)	-0.0653 (-0.62)
Monthly main DI benefits in 1,000 CHF in $t-1$	0.0810 (1.24)	0.0674 (0.82)
Monthly main DI benefits in 1,000 CHF in $t-2$	-0.125*** (-2.60)	-0.138** (-2.23)

Monthly main DI benefits in 1,000 CHF in $t-3$	0.0569 (1.34)	0.0628 (1.14)
Monthly child DI benefits in 1,000 CHF in $t-1$	-0.110 (-1.47)	-0.0698 (-0.78)
DI benefit entitlement in %	-0.0015 (-1.31)	-0.0033** (-2.42)
Civil status: married[#] (ref.: not married)	-0.0591 (-1.41)	-0.0202 (-0.39)
Helplessness allowance [#] (ref.: no)	0.367*** (2.69)	0.289 (1.34)
Supplementary benefit[#] (ref.: no)	0.146*** (2.79)	0.111* (1.67)
Supplementary benefit per case, amount in 1'000 CHF	-0.109*** (-3.75)	-0.093*** (-2.58)
Number of child DI benefits	0.0826* (1.94)	0.0606 (1.20)
Extraordinary DI benefit[#] (ref.: no)	-0.140* (-1.94)	-0.0729 (-0.81)
Nationality[#] (ref.: Swiss)		
Foreigner: German or Austria[#]	0.161** (2.13)	0.140 (1.46)
Foreigner: EU or EFTA countries[#]	-0.0270 (-0.57)	-0.0114 (-0.20)
Foreigner: rest of the world[#]	-0.0582 (-1.38)	-0.0192 (-0.37)
Daily allowance of the DI in $t-1$[#] (ref.: no)	-0.0540 (-0.49)	0.0155 (0.11)
Daily allowance of the unemployment insurance in $t-1$[#] (ref.: no)	0.577*** (9.29)	0.495*** (6.13)
Monthly income from paid employment in 1,000 CHF in $t-1$	-0.065*** (-3.21)	-0.046* (-1.82)
Monthly income from paid employment in 1,000 CHF in $t-2$	-0.017 (-0.79)	-0.021 (-0.86)
Monthly income from paid employment in 1,000 CHF in $t-3$	0.0131 (1.42)	-0.00280 (-0.17)
Income from paid employment in $t-1$[#] (ref.: no)	0.0729* (1.67)	0.00686 (0.12)
Income from paid employment in $t-2$[#] (ref.: no)	0.100*** (2.61)	0.167*** (3.58)
Income from paid employment in $t-3$[#] (ref.: no)	0.0589 (1.43)	0.0241 (0.44)
Participation in VR measures in $t-1$[#] (ref.: no)	0.909*** (5.89)	0.540*** (5.27)
Participation in VR measures in $t-2$[#] (ref.: no)	0.218** (2.38)	0.225* (1.92)
Participation in VR measures in $t-3$[#] (ref.: no)	0.284*** (3.37)	0.294*** (2.83)
Functional disorder[#] (ref.: others)		
Impairment of general condition[#]	0.0721 (1.38)	0.146** (2.19)
Behavioural disorders[#]	0.0842 (1.33)	0.203** (2.53)
Multiple mental disorders[#]	0.0403 (0.53)	0.135 (1.40)
At the trunc[#]	0.160* (1.82)	0.156 (1.41)
Multiple mental and physical disorders[#]	0.0652 (0.76)	0.140 (1.31)
Year dummies[#] (ref.: 2009)		
2010[#]	0.984*** (17.71)	0.981*** (17.32)
2011[#]	0.755***	

	(13.30)	
Constant	-3.281***	-3.819***
	(-11.11)	(-9.77)
Number of observations N	41,618	27,719
Observation participants N_1	908	579
Observation non-participants N_0	40,710	27,140
Pseudo R^2	0.143	0.131
Value of log likelihood	-3747.42	-2444.12
p-value likelihood ratio test	0.0000	0.0000
Average observed participation probability	2.178%	2.087%
Average predicted participation probability	2.182%	2.089%

Notes:[#]Dummy variable. t statistics in parentheses. $^*\, p < 0.10$, $^{**}\, p < 0.05$, $^{***}\, p < 0.01$

After the estimation of the propensity score the common support condition is analyzed. A graphical representation for both samples can be found in Figure A1 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.[6] 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.

B. Match-Quality

1. Balancing of Covariates

Balancing tests are based on the property $X \perp C | \hat{e}(X)$: 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 (2013) provides an overview. Some of the

---

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

procedures implemented by Leuven and Sianesi (2003) in the STATA command "pstest" are applied here.

Table 6 shows a detailed analysis separately by every covariate included in the propensity score. To keep it simple the detailed results are presented for kernel-based matching only. The corresponding results for radius matching can be found in Table A1 in the Appendix. Table 6 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 ($\bar{x}_1$) are shown. 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 6) being significantly reduced in the matched samples in comparison to the unmatched samples. This is also the main insight from Figure 2, which presents the standardized differences of all condition variables in histograms. There are, however, 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 (1985) designate a standardized difference of greater than 20% as large. Caliendo and Kopeinig (2008) argue 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 6 shows a measure proposed by Rubin (2001). 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 closer to 1.0 in the matched sample than in the unmatched sample.

The next step is to look at summary measures of the overall (im)balance of all conditioning variables. These are shown in Table 7 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 7 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 7 two summary measures proposed by Rubin (2001) 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 (2001) specifies 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 7 as well as in the histogram of the standardized differences in Figure 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.

TABLE 6 —MATCH QUALITY WITH RESPECT TO CONDITIONING VARIABLES – KERNEL-BASED MATCHING

		Program starters 2009–2011					Program starters 2009–2010				
		$\bar{x}_1$	$\bar{x}_0$	Std. diff. %	p-val. t-test	Var.-ratio	$\bar{x}_1$	$\bar{x}_0$	Std. diff. %	p-val. t-test	Var.-ratio
Years receiving DI (since 2000)	U	5.3	6.5	-33.4	0.000	0.94	5.4	6.0	-16.2	0.000	0.94
	M	5.3	5.2	2.5	0.600	1.03	5.4	5.4	1.8	0.771	0.93
Age	U	43.7	46.9	-34.3	0.000	1.09	43.7	46.4	-29.2	0.000	0.94
	M	43.7	43.4	3.1	0.520	1.08	43.7	43.6	1.3	0.821	1.04
Woman# (ref.: man)	U	47.0	47.4	-0.7	0.831	1.00	46.5	47.4	-1.9	0.659	1.00
	M	47.0	47.8	-1.6	0.727	1.00	46.5	47.4	-1.6	0.781	1.00
Type of diseases (ref.: congenital defects)											
mental#	U	12.4	7.3	17.3	0.000	1.52*	11.1	10.0	3.6	0.387	1.07
	M	12.4	12.1	1.1	0.837	0.98	11.1	10.9	0.5	0.939	1.01
nervous system #	U	55.7	54.9	1.8	0.601	1.00	57.2	53.3	7.7	0.067	0.99
	M	55.7	55.7	0.0	0.994	1.00	57.3	57.1	0.3	0.965	1.00
musculo-skeletal #	U	4.8	5.7	-3.8	0.275	0.85	4.5	5.4	-4.3	0.326	0.82
	M	4.8	5.1	-1.0	0.824	0.96	4.5	5.0	-2.3	0.696	0.93
other#	U	15.1	15.8	-2.0	0.549	0.96	15.0	15.4	-1.1	0.801	0.98
	M	15.1	14.9	0.5	0.917	1.01	15.1	14.9	0.5	0.934	1.00
injuries#	U	6.1	8.3	-8.7	0.015	0.75*	5.9	8.0	-8.3	0.063	0.75*
	M	6.1	6.0	0.3	0.946	1.01	5.7	5.8	-0.3	0.961	0.99
Monthly main DI benefits in 1,000 CHF in t-1	U	1217.4	1369.1	-23.7	0.000	1.11	1219.8	1332.0	-17.3	0.000	0.96
	M	1217.4	1208.2	1.4	0.769	1.06	1221.3	1217.2	0.6	0.916	1.00
Monthly main DI benefits in 1,000 CHF in t-2	U	1052.8	1283.6	-32.1	0.000	1.12	1060.6	1234.5	-23.9	0.000	0.99
	M	1052.8	1042.4	1.4	0.768	1.07	1062.4	1053.2	1.3	0.833	1.02
Monthly main DI benefits in 1,000 CHF in t-3	U	954.7	1188.6	-30.6	0.000	1.05	980.9	1141.7	-21.0	0.000	0.96
	M	954.7	943.5	1.5	0.760	1.06	982.6	973.3	1.2	0.839	1.02
Monthly child DI benefits in 1,000 CHF in t-1	U	208.9	222.9	-2.9	0.396	0.88	228.4	225.2	0.7	0.879	0.90
	M	208.9	219.3	-2.2	0.635	0.97	228.6	239.7	-2.3	0.699	0.97
DI benefit entitlement in %	U	72.6	81.9	-28.8	0.000	1.19	73.0	79.7	-20.3	0.000	0.95
	M	72.6	72.0	1.8	0.715	1.03	73.1	72.7	0.9	0.877	0.96
Civil status: married# (ref.: not married)	U	29.2	36.4	-15.5	0.000	0.89	32.5	35.5	-6.4	0.129	0.96
	M	29.2	29.5	-0.7	0.884	0.98	32.4	32.5	-0.4	0.945	1.00
Helplessness allowance # (ref.: no)	U	1.5	0.8	7.0	0.012	1.94*	0.9	0.6	3.1	0.420	1.43*
	M	1.5	1.5	0.7	0.901	1.04	0.9	0.8	0.4	0.954	1.03
Supplementary benefit# (ref.: no)	U	34.5	35.2	-1.5	0.661	0.99	33.2	33.7	-1.1	0.790	0.99
	M	34.5	34.5	-0.2	0.974	1.00	33.2	33.0	0.5	0.936	1.01
Supplementary benefit per case in 1'000 CHF	U	436.2	526.75	-10.5	0.004	0.64*	430.3	495.4	-7.6	0.091	0.75*
	M	436.2	430.18	0.7	0.867	1.00	431.1	423.1	0.9	0.863	1.04
Number of child DI benefits	U	0.427	0.404	2.8	0.399	1.12	0.465	0.407	6.7	0.097	1.19
	M	0.427	0.456	-3.4	0.482	0.95	0.464	0.487	-2.7	0.657	1.10
Extraordinary DI benefits # (ref.: no)	U	5.8	6.4	-2.2	0.529	0.92	5.5	6.1	-2.4	0.574	0.91
	M	5.8	5.7	0.6	0.895	1.02	5.5	5.7	-0.6	0.923	0.97
Nationality# (ref.: Swiss)											
Foreigner: German or Austria#	U	5.1	3.1	9.7	0.001	1.57*	4.7	3.1	7.8	0.040	1.45*

	M	5.1	4.8	1.2	0.813	1.05	4.7	4.6	0.5	0.934	1.02
Foreigner: EU or EFTA countries#	U	13.7	13.7	-0.2	0.945	1.00	14.2	13.7	1.2	0.768	1.03
	M	13.7	13.4	0.9	0.849	1.02	14.2	14.0	0.4	0.946	1.01
Foreigner: rest of the world#	U	23.0	25.4	-5.5	0.107	0.93	25.2	25.4	-0.4	0.932	1.00
	M	23.0	23.8	-1.8	0.702	0.95	25.1	25.4	-0.8	0.896	1.00
Daily allowance of the DI in t-1# (ref.: no)	U	3.6	1.2	16.0	0.000	2.90**	3.6	1.6	12.4	0.000	2.09**
	M	3.6	3.7	-0.3	0.955	0.99	3.6	3.2	2.9	0.669	1.08
Daily allowance of the unemployment insurance in t-1# (ref.: no)	U	13.1	2.3	41.1	0.000	3.74**	10.5	2.6	32.6	0.000	2.93**
	M	13.1	12.6	1.9	0.754	0.98	10.4	10.7	-1.3	0.857	1.03
Monthly income from paid employment in 1,000 CHF in t-1	U	748.8	757.4	-0.6	0.868	0.64*	766.6	801.2	-2.4	0.608	0.60*
	M	748.8	763.0	-1.0	0.796	1.23	765.8	801.3	-2.5	0.612	1.29*
Monthly income from paid employment in 1,000 CHF in t-2	U	917.2	865.3	3.3	0.364	0.71*	828.1	878.6	-3.2	0.486	0.66*
	M	917.2	899.9	1.1	0.797	1.05	827.4	853.5	-1.7	0.746	1.11
Monthly income from paid employment in 1,000 CHF in t-3	U	1212.5	916.8	15.6	0.000	0.92	1089.6	993.3	5.1	0.261	0.69*
	M	1212.5	1205.7	0.4	0.951	0.46**	1084.7	1108.1	-1.2	0.816	1.02
Income from paid employment in t-1# (ref.: no)	U	68.3	55.7	26.2	0.000	0.93	75.1	62.1	28.4	0.000	0.87
	M	68.3	69.2	-1.8	0.683	1.02	75.1	75.7	-1.3	0.820	1.01
Income from paid employment in t-2# (ref.: no)	U	67.4	66.8	1.2	0.724	0.98	59.6	61.7	-4.3	0.302	1.02
	M	67.4	67.1	0.6	0.905	0.99	59.5	60.4	-1.8	0.760	1.01
Income from paid employment in t-3# (ref.: no)	U	72.8	57.7	32.2	0.000	0.85	76.9	63.0	30.6	0.000	0.83
	M	72.8	73.2	-0.9	0.831	1.02	76.8	77.3	-1.0	0.851	1.02
Participation in VR measures in t-1# (ref.: no)	U	16.0	2.0	50.6	0.000	4.66**	6.4	1.3	26.8	0.000	4.01**
	M	16.0	16.0	0.0	0.997	1.00	6.2	5.8	2.2	0.768	1.03
Participation in VR measures in t-2# (ref.: no)	U	4.4	1.5	17.3	0.000	2.81**	4.0	1.6	14.3	0.000	2.34**
	M	4.4	4.3	0.8	0.893	1.02	3.8	4.1	-1.8	0.802	0.96
Participation in VR measures in t-3# (ref.: no)	U	5.3	1.7	19.4	0.000	2.87**	5.4	1.9	18.3	0.000	2.57**
	M	5.3	5.2	0.4	0.947	1.01	5.2	5.1	0.5	0.940	1.01
Functional disorder# (ref.: others)											
impairment of general condition#	U	40.6	40.9	-0.4	0.898	1.00	40.8	39.6	2.5	0.557	1.01
	M	40.6	40.7	-0.1	0.983	1.00	40.7	40.8	-0.4	0.948	1.01
behavioral disorders#	U	21.4	22.0	-1.6	0.633	0.97	23.7	21.6	5.0	0.224	1.08
	M	21.4	21.7	-0.8	0.862	0.97	23.7	23.8	-0.3	0.964	0.99
multiple mental disorders#	U	8.0	7.8	1.0	0.774	1.03	7.9	7.5	1.8	0.669	1.06
	M	8.0	7.9	0.7	0.884	1.02	8.0	7.8	0.5	0.930	1.02
at the trunc#	U	4.4	4.8	-1.8	0.600	0.92	4.3	4.7	-1.8	0.672	0.92
	M	4.4	4.6	-0.7	0.872	0.96	4.3	4.3	0.0	0.996	1.00
multiple mental and physical disorders#	U	4.2	5.0	-3.8	0.277	0.85	4.3	4.9	-2.6	0.548	0.89
	M	4.2	4.1	0.2	0.968	1.01	4.3	4.4	-0.2	0.973	0.99
Year dummies# (ref.: 2009)											
2010#	U	58.0	33.3	51.2	0.000	1.62*	91.0	50.0	100.7	0.000	0.77*
	M	58.0	59.9	-3.9	0.409	0.96	91.0	90.1	2.2	0.607	0.80
2011#	U	36.2	33.3	6.1	0.067	1.05					
	M	36.2	32.6	7.6	0.106	0.98					

Notes: #Dummy variable. Standardized difference $\frac{|\bar{x}_1 - \bar{x}_0|}{\sqrt{(V(x_1)+V(x_0))/2}}$ (Rosenbaum and Rubin, 1985). Ratio of the variance of the residuals orthogonal to the linear index of the propensity score in the treated group over the non-treated group (Rubin, 2001). * if "of concern", i.e. variance ratio in [0.5, 0.8) or (1.25, 2]; ** if "bad", i.e. variance ratio <0.5 or >2.

TABLE 7 —MATCH QUALITY, SUMMARY MEASURES

| Matching method		Pseudo-R^2	p-value Likeli-hood 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
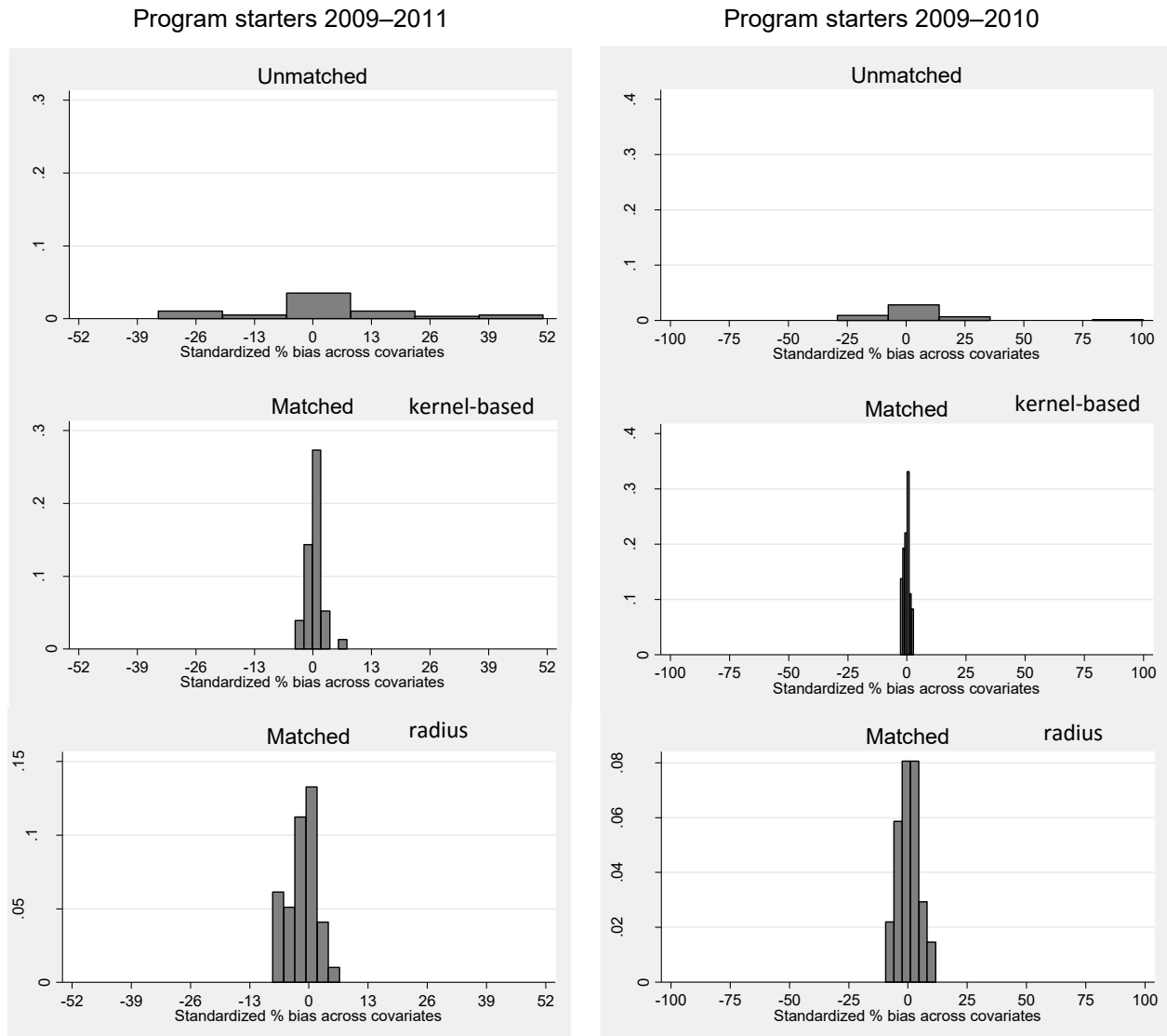
FIGURE 2— HISTOGRAM OF STANDARDIZED DIFFERENCES OF ALL CONDITIONING VARIABLES IN TABLE 6

2. Pre-Program Outcome Variable

The "pre-program test" proposed by Heckman and Hotz (1989) 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" (see previous subsection). 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). Time means here 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 to 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 on average lower benefits. 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 Figure 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 Figure 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 %.

In the next subsection the ATT effects, which means, the differences between the treated and the control group in the years after the individual starting year, are analyzed in greater detail.
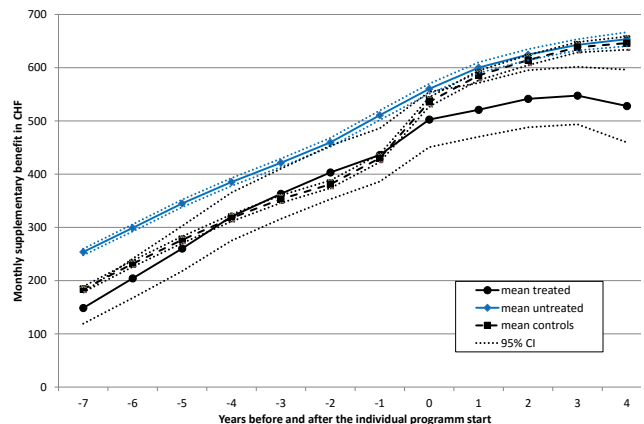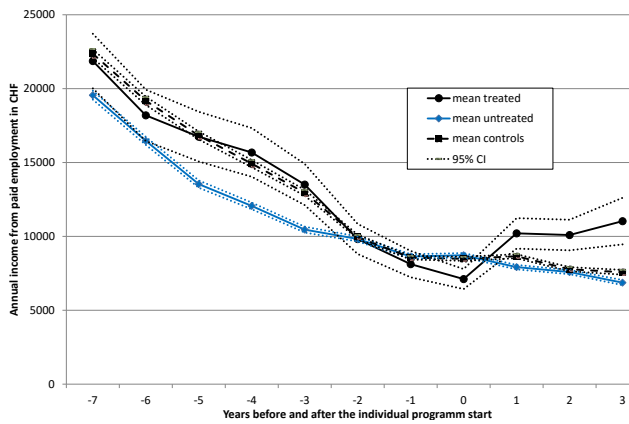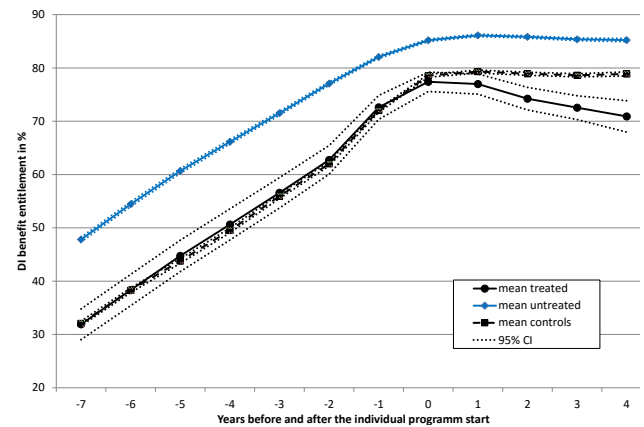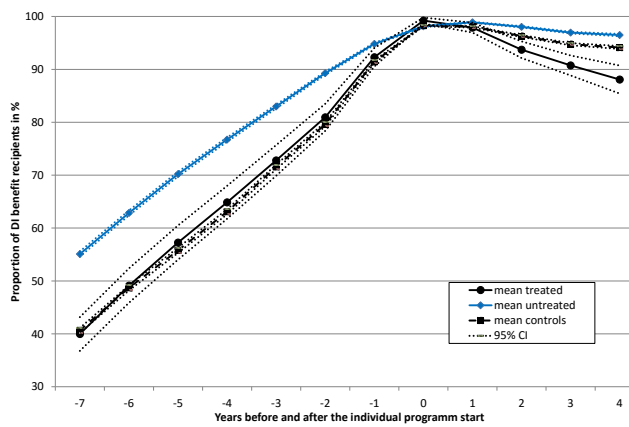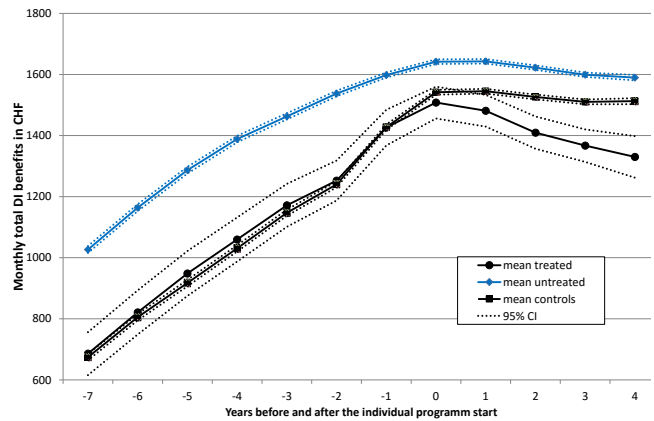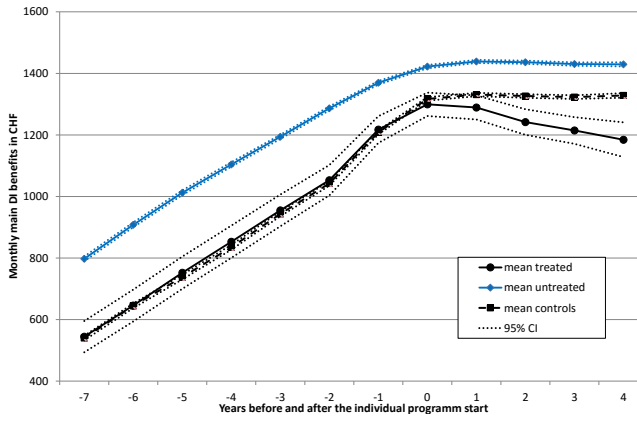
Notes: All variables are prices adjusted to the base year 2009. Persons with no DI benefits are included with a benefit of zero. Also persons with no income from paid employment or supplementary benefit are included with a value of zero. See Table 4 for definition of the outcome variables.

FIGURE 3 — MEAN OUTCOME VARIABLES OF THE TREATED, UNTREATED AND CONTROL GROUP

C. Estimated ATT-Effects

Table 8 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 9 presents the corresponding ATT effects on income variables. As explained in Section IV 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 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 percentage points lower in the second year after individual starting year ($t + 2$), 4.9 percentage point in $t + 3$ and 7.8 percentage points in $t + 3$ (Table 8). In the group of participants in $t + 2$, the proportion of persons earning income from paid employment is 13 percentage points higher than in the control group. As a result, the average annual income from paid employment in $t + 2$ is CHF 2,750 higher, corresponding to a relative increase of approximately 37%.[7] In $t + 3$ the ATT effect on income is CHF 3,975 (approx. +56%). The amount of the monthly main DI benefit is reduced by 3.8% in $t + 1$ compared with 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 1,752, 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 2,112.

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 percentage points in $t + 1$ to -8.3 percentage points in $t + 4$. The monthly supplementary benefits are reduced by CHF 71 (=12%) in $t + 1$ up to 125 (=19%) in $t + 4$.

---

[7] This relative increase of 37% can be roughly calculated as follows: from the graph in the bottom-left of Figure 3 it can be seen that the counterfactual income is approximately CHF 7,500. CHF 2,750 / CHF 7,500 is approximately 37%.

TABLE 8 — ESTIMATED ATT-EFFECTS ON DI BENEFITS AND SB – KERNEL-BASED MATCHING, T-STATS IN PARENTHESES

Outcome variable	Matching method	$t+1$ Starters 2009–2011 Outcome 2010–2012	$t+2$ Starters 2009–2011 Outcome 2011–2013	$t+3$ Starters 2009–2011 Outcome 2012–2014	$t+4$ 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

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

TABLE 9 — ESTIMATED ATT-EFFECTS ON INCOME – t-STATS IN PARENTHESES

Outcome Variable	Matching method	$t+1$ Starters 2009–2011 Outcome 2010–2012	$t+2$ Starters 2009–2011 Outcome 2011–2013	$t+3$ 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	1,549** (2.83)	2,313*** (4.16)	3,479*** (4.85)
	Kernel, DiD	1,983*** (3.85)	2,747*** (4.95)	3,975*** (5.37)
	Radius, DiD	2,198*** (3.64)	2,922*** (4.57)	4,017*** (4.75)
Number of participants		908	908	578
Number of controls		40,710	40,710	27,140

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

VII. Sensitivity Analysis

A. Unobserved Confounders

Since the analysis is based on the selection on observables assumption / CIA, the question arises whether and under which conditions unobserved variables may significantly change the estimated ATT effects.[8] Possible candidates are variables relating to second and third pillar of the social security system (see Section II), educational background and personality traits.

For this purpose, the simulation-based sensitivity analysis proposed by Ichino et al. (2008) and implemented by Nannicini (2007) in STATA is used to check the robustness of the results with respect to possible deviations from the CIA. The idea is to simulate a potential binary confounding variable, U (an unobserved dummy variable which is not included in observed conditioning variables X), using different distributional assumptions, and to check whether the inclusion of this variable affects the estimated ATT-effect (see Boockmann and Hagen 2008).

For a binary treatment, C, and a binary outcome, Y, the distribution of the binary variable, U, is characterized by four parameters:

---

[8] Since the matching approach applied conditions on pre-treatment outcome variables and additionally uses DiD, the question can be reformulated: are there *time-varying* unobserved variables which may bias the results?

$$p_{kl} \equiv Pr(U = 1|C = k, Y = l, X) = p_{kl} \equiv Pr(U = 1|C = k, Y = l)$$

with $k, l \in \{0,1\}$ and the assumption of conditional independence of U with respect to X. Given these parameters, for each person a particular value of U is drawn and included in the estimation of the propensity score and the calculation of the ATT effect. This procedure is repeated many times and the estimated ATT effects of each draw are averaged. The resulting "averaged simulated ATT effect" is an estimate of the ATT effect which is robust to the inclusion of the unobserved variable U characterized by the distributional assumption in p_{kl} (Boockmann and Hagen 2008). In order to simulate realistic and relevant confounding variables, Ichino et al. (2008) propose to choose the parameters p_{kl} such that the distribution of U is similar to the distribution of existing binary covariates in the condition set X. If the problem is overestimation of the ATT effect, the interesting potential confounders are those that have either a positive or a negative effect both on the untreated outcome and selection into treatment. For underestimation, the relevant case is that the effect on the outcome is positive (negative) and the selection effect is negative (positive). In order to reveal the kind of failure of the CIA implied by the simulated confounder U, the estimated effect of U on the selection into treatment (selection effect) and the estimated effect of U on the outcome of untreated subjects (outcome effect) are computed as odds ratios (see Ichino et al. 2008). This enables the researcher to assess how strong selection and outcome effects must be to exert a sizable impact on the ATT effect.

Since the simulation-based sensitivity analysis is constructed for binary outcome variables and since it requires a lot of computing time (several weeks) the analysis is restricted to the binary outcome variable *DI benefit (yes)* in $t + 3$, as an example. We set the number of iterations (repetitions for the procedure) to 250. Again due to the computing time, it is not possible to bootstrap the standard errors and hence no standard errors are reported. The simulation is based on the kernel-based matching, i.e., the preferred specification found in Section V.B.

The first row of Table 10 shows the baseline estimate (-4.0 percentage point) without any confounders, which corresponds to the results in Table 8 (row "DI benefit (yes) in %", "Kernel, Levels" and column $t + 3$). The second row shows ATT effects with a neutral confounder, such that the average effect on the untreated outcome and on selection is zero. The other rows show the simulated ATT effects if the distribution of U resembles the distribution of important binary variables contained in X. The deviations from the ATT effect of -4.0 are very small. All in all, the results of the matching estimator seem to be robust against the failure of the CIA due to neglected

unobserved determinants of outcome and selection, although it should be kept in mind, that it was not possible to compute standard errors and that the whole sensitivity analysis presents in this section was carried out only in an exemplary manner for one outcome variable and with some of the conditioning variables.

TABLE 10 — SENSITIVITY ANALYSIS: EFFECT OF CALIBRATED CONFOUNDERS FOR KERNEL MATCHING ESTIMATES OF THE ATT EFFECTS ON THE OUTCOME VARIABLE *DI BENEFIT (YES)* IN $t + 3$

	Fraction $U = 1$				Outcome effect	Selection effect	ATT effect in %
	p_{11}	p_{10}	p_{01}	p_{00}			
No confounder	0.0	0.0	0.0	0.0	–	–	-4.0
Neutral confounder	0.5	0.5	0.5	0.5	1.008	0.992	-4.0
Confounder, such as:							
Woman	0.48	0.42	0.47	0.51	0.853	0.983	-4.0
Type of diseases: congenital defects	0.12	0.12	0.12	0.12	0.749	1.798	-3.9
Married	0.28	0.38	0.36	0.48	0.603	0.711	-4.2
Supplementary benefit	0.35	0.29	0.36	0.22	2.003	0.982	-4.0
Nationality: Swiss	0.05	0.05	0.03	0.04	0.905	1.651	-4.0
Income from paid employment	0.68	0.71	0.56	0.58	0.906	1.710	-3.9
Participation in VR measures	0.04	0.06	0.01	0.03	0.447	3.022	-3.9
Functional disorder: impairment of general condition	0.41	0.42	0.41	0.43	0.908	0.989	-4.0
Year 2010	0.58	0.62	0.33	0.32	1.064	2.796	-4.0

Notes: The simulated ATT effects are based on 250 repetitions.

B. Bandwidth Choice

The bandwidth choice does not seem to be very important here. This is shown for the example of the outcome variable *monthly total DI benefit in CHF (main + child's benefit)* estimated by kernel-based matching combined with DiD. The results are presented in Table 11, where for different bandwidths the ATT effects as well as summary measures for the match quality are shown. Two conclusions can be drawn from this overview: (1.) with respect to the summary measures for the match quality a bandwidth of 0.06 (suggested by Silverman's rule of thumb) seems to be a "good choice". However, a bandwidth of 0.1 does not to seem worse. (2.) the differences in the estimated ATT effects are very small. Part of the differences may by be driven by treated persons being of "off support" in case of the smaller bandwidth (0.01).

Hence, the estimated ATT effects are not very sensitive with regard to the choice of the bandwidth in the case of the kernel-based matching.

C. Common Support Condition

In Section VI the common support condition is imposed before matching by dropping treated individuals whose propensity score is higher than the maximum than the propensity score of the potential controls.[9] However, Figure A1 in the Appendix indicates that the number of potential controls may be rather low in the upper tails of the density. Hence, an alternative method of imposing common support is used here: a certain proportion of the treated individuals for which the propensity score density of the untreated observations is the lowest is dropped. Table 12 shows the results for dropping 1%, 2%, 5% and 10% of treated individuals for ATT effect on *monthly total DI benefit* and corresponding summary measures of the match quality. The obvious problem with this approach is that the sample of treated persons becomes smaller and smaller. This implies that, if the treatment effect is heterogeneous inside and outside the support, the estimated ATT effects also change with the sample of treated persons (Lechner 2008). Hence, it is unclear whether changing ATT effects indicate a problem with the common support or simply heterogeneity of the treatment effects. On the other hand, if the ATT effects do not change much, this is an indication that the results are robust with respect to the way the common support is imposed.

Looking at Table 12 and comparing it with the "preferred specification" (bandwidth 0.06) in Table 11 the following insights can be gained: (1.) the indicators for the match quality improve only slightly and this is not unambiguous; (2.) the results change only a little and not systematically with respect to the amount of trimming. Summarizing, one can conclude that a substantial bias from a lack of common support is unlikely.

---

[9] Furthermore, in the matching procedure support is implicitly imposed by the Epanechnikow kernel in the case of kernel-based matching and the radius in the case of radius matching.

TABLE 11 —SENSITIVITY TO BANDWIDTH CHOICE, PROGRAM STARTERS 2009-2011

| Band-width | Treated: off support | ATT effect on monthly total DI benefit in CHF (t-stat.) | | | Match quality, summary measures | | | | | |
		$t+1$	$t+2$	$t+3$	Pseudo-R^2	Likelihood ratio test	Mean bias	Med. bias	Rubin's B	Rubin's R
0.01	5	-56 (-3.07)	-108 (-4.86)	-135 (-5.28)	0.003	9.63	1.7	1.5	13.8	1.03
0.03	1	-62 (-3.58)	-115 (-5.38)	-142 (-6.04)	0.003	7.49	1.5	1.1	12.9	0.97
*0.06	0	-63 (-3.88)	-115 (-5.68)	142 (-6.12)	0.003	6.90	1.3	0.9	12.3	0.99
0.10	0	-62 (-4.06)	-115 (-6.18)	-142 (-6.88)	0.003	6.59	1.3	1.0	12.1	1.02

Notes: All estimated ATT effects are based on kernel-based matching with the Epanechnikow kernel and DiD. The matched sample is based on 908 treated minus individuals being "off support" as well as 40,710 controls. The bandwidth of 0.06 is suggested by Silverman's rule and this is the preferred specification used for the evaluation.

TABLE 12 —SENSITIVITY TO TRIMMING THE COMMON SUPPORT, PROGRAM STARTERS 2009-2011

| Trim | Treated: off support | ATT effect on monthly total DI benefit in CHF (t-stat.) | | | Match quality, summary measures | | | | | |
		$t+1$	$t+2$	$t+3$	Pseudo-R^2	Likelihood ratio test	Mean bias	Med. bias	Rubin's B	Rubin's R
1%	9	-59 (-3.66)	-112 (-5.42)	-137 (-5.85)	0.003	7.15	1.4	1.0	12.6	1.01
2%	18	-64 (-3.83)	-118 (-5.81)	-145 (-6.40)	0.003	6.29	1.2	0.9	11.9	0.97
5%	45	-59 (-3.72)	-112 (-5.77)	-139 (-6.27)	0.003	6.24	1.3	1.0	12.0	0.98
10%	90	-70 (-4.09)	-118 (-5.71)	-148 (-6.41)	0.002	4.95	1.0	0.8	11.0	0.92

Notes: All estimated ATT effects are based on kernel-based matching with the Epanechnikow kernel and DiD. The matched sample is based on 908 treated minus individuals being "off support" as well as 40,710 controls.

VIII. 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 a rough guideline only. 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 8,819 (= EUR 8038 or USD 8,970)[10] per participant were 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 four 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 Section IV: (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 applying 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 reached 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.

- **S1** is the most optimistic scenario: the ATT effect in $t + 1$ equals the estimates in Table 8. From $t + 5$ until old-age pension the ATT effects are constant at the level of the weighted average of the periods $t + 2$, $t + 3$ and $t + 4$, with the weights depending on the number of the participants (908, 908, 578).

---

[10] Exchange rates at 16-March-2016.

- **S2**: the ATT effects are from $t + 1$ to $t + 4$ according to the estimates in Table 8. From $t + 5$ until old-age pension the ATT effects are constant at the level of the period $t + 2$.

- **S3**: the ATT effects are in $t + 1$ to $t + 3$ according the estimates in Table 8. From $t + 4$ to $t + 10$ they are at the level of period $t + 3$. From $t + 11$ on the ATT effects are assumed to be zero.

- **S4** is the most pessimistic scenario: the ATT effect is at its maximum in $t + 4$ according to the estimates (Table 8). The ATT effect of $t + 5$ is set to the value of $t + 3$; the ATT effect of $t + 6$ is set to the value of $t + 2$, and the ATT effect of $t + 7$ is set to $t + 1$. From $t + 8$ the effects are assumed to be zero. Hence, the effectiveness is assumed to be temporary for seven years.
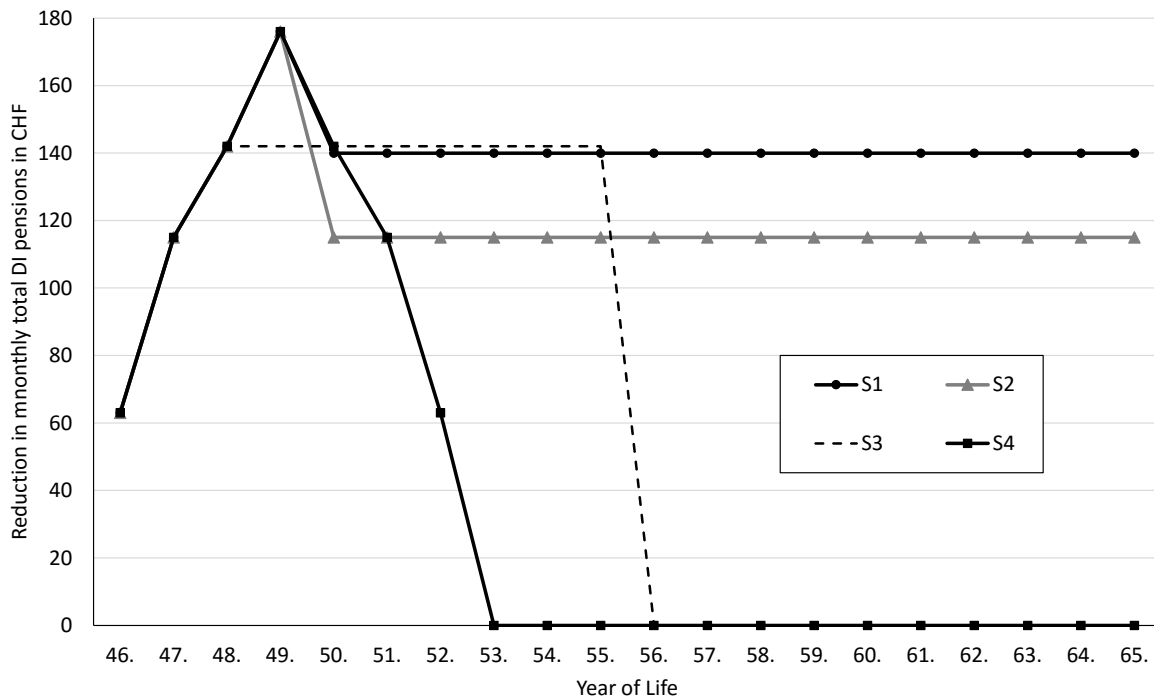
Notes: $t + 1$ corresponds to the 46th year of life. The first four years (46., 47., 48., 49.) correspond to the kernel-based DiD matching estimates of the ATT effect in Table 8.

FIGURE 4—GRAPHICAL REPRESENTATION OF THE SCENARIOS ASSUMED FOR THE COST BENEFIT ANALYSIS

Furthermore, since the ATT effects on total taxable income[11] are estimated, it is possible to calculate the additional social security contributions resulting from higher income. The total

[11] Table 9 presents only the ATT effects on income from paid employment. The effects on total income are very similar (t-stat.): CHF 1,932 (3.73) in $t + 1$, CHF 2,521 (4.33) in $t + 2$, and CHF 3922 (4.91) in $t + 3$.

contribution rate, which is assumed be constant in the future, is 12.5%.[12] However, due to issues involving data availability, as described in Section IV, the ATT effects are only estimated up to $t + 3$. Hence, the scenarios described above have to be adjusted.[13] While it is possible to estimate the effects on social security contribution, this is not possible with regard to income tax (or other taxes), because not all necessary information (such as further incomes) is available.

TABLE 13—CURRENT VALUES OF EXPECTED REDUCTIONS IN TOTAL DI BENEFIT PAYMENTS AND ADDITIONAL SOCIAL SECURITY CONTRIBUTIONS BY SCENARIO AND DISCOUNT RATE (PER PARTICIPANT IN CHF)

	Real discount rate		
Scenario	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	6,800	6,200	5,600
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	4,700	4,300	3,900
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	9,900	9,400
Increase in social security contributions	3,400	3,200	3,100
Sum	28,800	27,200	25,800
S4			
Decrease in total DI benefits	9,400	9,100	8,700
Decrease in supplementary benefits	7,100	6,900	6,600
Increase in social security contributions	1,600	1,500	1,500
Sum	18,100	17,500	16,800

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 (2015) indicate for Switzerland an average of

[12] 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

[13] S1: the ATT effects are from $t + 4$ until old-age pension constant at the level of weighted average of the periods $t + 2$ and $t + 3$, with the weights depending on the number of the participants. S2: the ATT effects are from $t + 4$ until old-age pension constant at the level of the period $t + 2$. S3: the ATT effects are in $t + 1$ to $t + 2$ according the estimates. From $t + 3$ to $t + 10$ they are at the level of period $t + 2$. From $t + 11$ on the ATT effects are assumed to be zero. S4: the ATT effect of $t + 4$ is set to the value of $t + 2$ and the ATT effect of $t + 5$ is set to the value of $t + 1$. From $t + 6$ the ATT effects are assumed to be zero.

2.8% for the period 1981 to 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. Furthermore, historical data may not provide any information regarding future real interest rates. Given the low returns on Swiss bonds and the discussion of "secular stagnation" (Baldwin and Teulings 2014), 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 contribution are calculated.

Table 13 shows the results, which are rounded to CHF 100 in order to avoid spurious accuracy. Comparing the social security system's investment of CHF 8,819 per participant with the current values of expected reductions in 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 8,000 and CHF 48,300 per participant.

IX. 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 specializes in this field. The project was additional to the already existing "traditional" VR programs. However, 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 from the Swiss social security system. For this purpose a rich administrative panel data set is analyzed. However, 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 difference-in-differences approach is applied in addition to matching. Moreover, the sensitivity of the results with regard to this problem is simulated. 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 war effective for the participants in that the DI benefit amounts paid to them could be lowered by raising their levels of paid employment and income. The DI benefits can be reduced by about 10%. Up to four 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 to 6.5 times. Subtracting the costs of the project, the expected net benefit of the social security system is between CHF 8,000 (= EUR 7,290 or USD 8,136) and CHF 48,300 (=EUR 44,016 or USD 49,121) per participant.[14]

How can these favorable results be explained? First of all, one should note, that in general job search assistance programs have been found to be relatively effective (see Section I). One explanation for this effectiveness and for the results of this paper is the absence of lock-in effect, which often hampers the effectiveness of training programs (Brown and Koettl, 2015). This is especially true for the coaching program evaluated here. Figure 3 in Section VI.B.2 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. 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 (2014) find that a strategy focusing on rapid placement in the regular labor market is superior to vocational training or sheltered employment.

---

[14] Exchange rates at 16-March-2016.

Given the fact that many participants had massive health problems and that they had received DI benefits for many years before starting the program, the conclusion that a pure placement coaching program is effective is remarkable. 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, financial incentives or threat effects.

References

Aakvik, A., Heckman, J. J., & Vytlacil, E. J. (2005). Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs. *Journal of Econometrics*, *125*(1), 15-51.

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, 47-57.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, *46*(3), 399-424.

Baldwin, R., & Teulings, C. (2014). Secular stagnation: facts, causes and cures. London: Centre for Economic Policy Research-CEPR.

Becker, S. O., & Caliendo, M. (2007). Sensitivity analysis for average treatment effects. The Stata Journal 7 (1), 71–83.

Biewen, M., B. Fitzenberger, A. Osikominu and M. Waller (2014). The Effectiveness of Public Sponsored Training Revisited: The Importance of Data and Methodological Choices, *Journal of Labor Economics, 32*(4), 837-897.

Boockmann, B., & Hagen, T. (2008). Fixed-term contracts as sorting mechanisms: Evidence from job durations in West Germany. *Labour Economics*, *15*(5), 984-1005.

Bound, J., & Burkhauser, R. V. (1999). Economic analysis of transfer programs targeted on people with disabilities. *Handbook of Labor Economics*, *3*, 3417-3528.

Brown, A. J., & Koettl, J. (2015). Active labor market programs-employment gain or fiscal drain?. *IZA Journal of Labor Economics*, *4*(1), 1-36.

Bütler, M., Deuchert, E., Lechner, M., Staubli, S., & Thiemann, P. (2015). Financial work incentives for disability benefit recipients: lessons from a randomised field experiment. *IZA Journal of Labor Policy*, *4*(1), 1-18.

Burkhauser, R. V., Daly, M. C., McVicar, D., & Wilkins, R. (2014). Disability benefit growth and disability reform in the US: lessons from other OECD nations. *IZA Journal of Labor Policy*, *3*(1), 1-30.

Card, D., Kluve, J., & Weber, A. (2010). Active labour market policy evaluations: a meta-analysis*. *The Economic Journal*, *120*(548), F452-F477.

Caliendo, M., Mahlstedt, R., & Mitnik, O. (2014). *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.

Campolieti, M., Gunderson, M. K., & Smith, J. A. (2014). The effect of vocational rehabilitation on the employment outcomes of disability insurance beneficiaries: new evidence from Canada. *IZA Journal of Labor Policy*, *3*(1), 1-29.

Dean, D., Pepper, J., Schmidt, R., & Stern, S. (2015). The effects of vocational rehabilitation for people with cognitive impairments. *International Economic Review*, *56*(2), 399-426.

Dehejia, R. (2005). Practical propensity score matching: a reply to Smith and Todd. *Journal of Econometrics*, *125*(1), 355-364.

Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association*, *94*(448), 1053-1062.

Fredriksson, P., & Johansson, P. (2008). Dynamic treatment assignment: the consequences for evaluations

using observational data. *Journal of Business & Economic Statistics*, *26*(4), 435-445.

Frölich, M., Heshmati, A., & Lechner, M. (2004). A microeconometric evaluation of rehabilitation of long-term sickness in Sweden. *Journal of Applied Econometrics*, *19*(3), 375-396.

Galdo, J. C., Smith, J., & Black, D. (2008). Bandwidth Selection and the Estimation of Treatment Effects with Unbalanced Data. *Annales d'Economie et de Statistique*, (91), 189-216.

Ichino, A., Mealli, F., & Nannicini, T. (2008). From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity?. *Journal of Applied Econometrics*, *23*(3), 305-327.

Imbens, G. W. (2015). Matching methods in practice: Three examples. *Journal of Human Resources*, *50*(2), 373-419.

Heckman, J. J., & Smith, J. A. (1999). The Pre-Programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies. *The Economic Journal*, *109*(457), 313-348.

Heckman, J. J., LaLonde, R. J., & Smith, J. A. (1999). The economics and econometrics of active labor market programs. *Handbook of Labor Economics*, *3*, 1865-2097.

Heckman, J. J., Ichimura, H., & Todd, P. (1998a). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, *65*(2), 261-294.

Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998b). Characterizing selection bias using experimental data. *Econometrica, 66*(5): 1017–1098.

Heshmati, A., Engström, L.G. (2001). *Estimating the effects of vocational rehabilitation programs in Sweden*, in: Lechner, M., Pfeiffer, F. (eds.), Econometric Evaluation of Labour Market Policies. ZEW Economics Studies 13. Physica-Verlag: New York, pp. 183-210.

Høgelund, J., & Holm, A. (2006). Case management interviews and the return to work of disabled employees. *Journal of Health Economics* 25(3), 500-519.

Huber, M., Lechner, M., & Steinmayr, A. (2015). Radius matching on the propensity score with bias adjustment: tuning parameters and finite sample behaviour. *Empirical Economics*, *49*(1), 1-31.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, 604-620.

Lechner, M. (2008). A Note on the Common Support Problem in Applied Evaluation Studies. *Annales d'Economie et de Statistique*, (91), 217-235.

Lechner, M., & Wunsch, C. (2013). Sensitivity of matching-based program evaluations to the availability of control variables. *Labour Economics*, *21*, 111-121.

Leuven, E. and B. Sianesi. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. http://ideas.repec.org/c/boc/bocode/s432001.html. This version 4.0.11.

Lechner, M., Miquel, R., & Wunsch, C. (2011). Long-run effects of public sector sponsored training in West Germany. *Journal of the European Economic Association*, *9*(4), 742-784.

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.

Millimet, D. L., & Tchernis, R. (2009). On the specification of propensity scores, with applications to the analysis of trade policies. *Journal of Business & Economic Statistics*, *27*(3), 397-415.

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

Nannicini, T. (2007). Simulation-based sensitivity analysis for matching estimators. *Stata Journal*, *7*(3), 334.

Rosenbaum, P. R. (2002). Observational Studies. 2nd ed. New York: Springer.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33-38.

Organisation for Economic Co-operation and Development. (2010). Sickness, Disability and Work: Breaking the Barriers: A Synthesis of Findings across OECD Countries. OECD Publishing.

Sianesi, B. (2004). An evaluation of the Swedish system of active labor market programs in the 1990s. *Review of Economics and Statistics*, *86*(1), 133-155.

Sianesi, B. (2008). Differential effects of active labour market programs for the unemployed. *Labour Economics*, *15*(3), 370-399.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall.

Smith, J. and Todd, P. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics* 125(1–2): 305–353.

Thomsen, S. (2009). *Job search assistance programs in Europe: evaluation methods and recent empirical findings,* FEMM Working Paper No. 18. Otto-von-Guericke University Magdeburg.

World Bank (2015). World Development Indicators, retrieved 09-Feb-2016.

Appendix

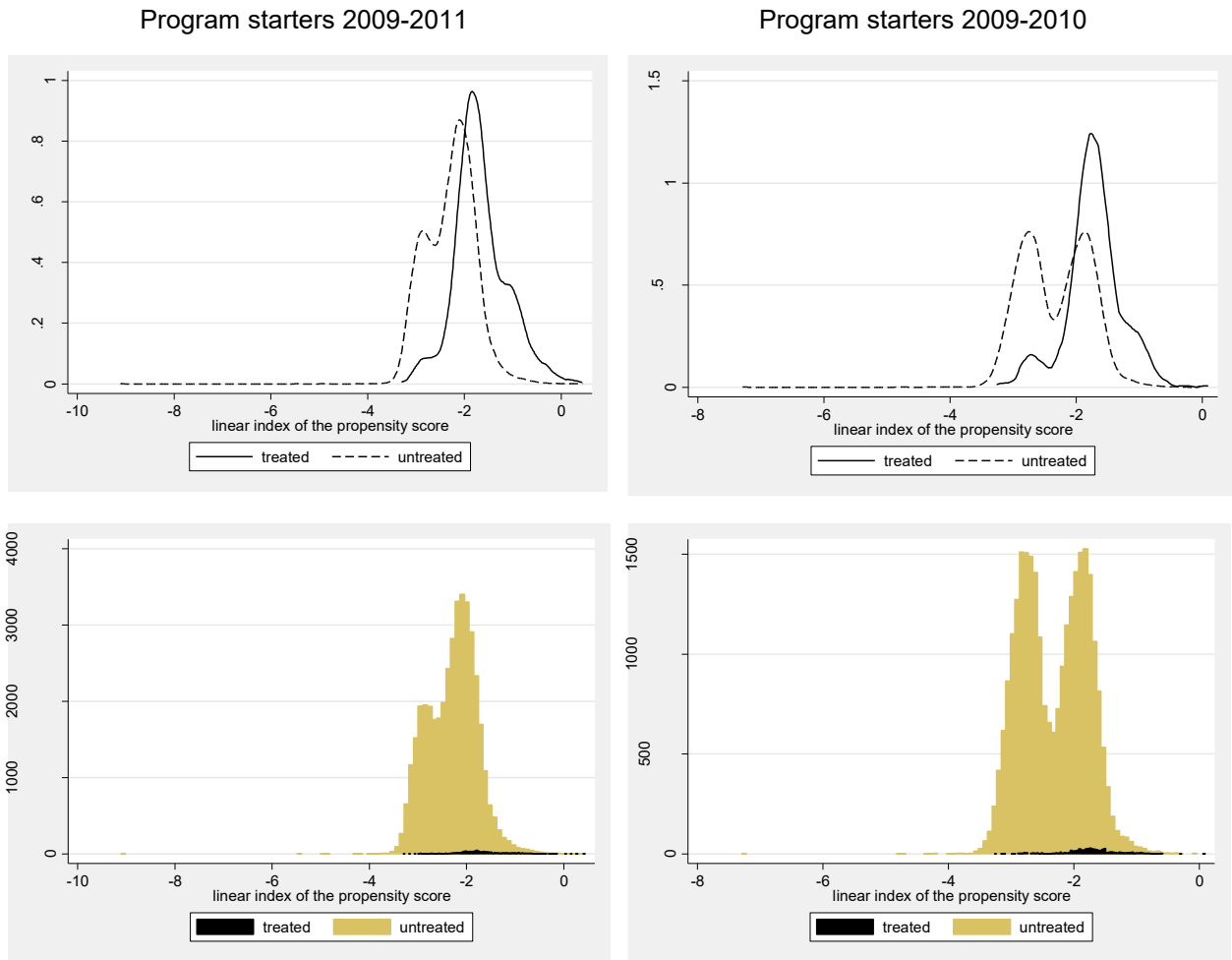Program starters 2009-2011

Program starters 2009-2010

FIGURE A1: KERNEL DENSITY ESTIMATE AND HISTOGRAM (FREQUENCY) OF THE PROPENSITY SCORE ESTIMATE IN TABLE 5

TABLE A1—MATCH QUALITY WITH RESPECT TO CONDITIONING VARIABLES – RADIUS MATCHING

		Program starters 2009-2011					Program starters 2009-2010				
		$\bar{x}_1$	$\bar{x}_0$	Std. diff. %	p-val. t-test	Var.-ratio	$\bar{x}_1$	$\bar{x}_0$	Std. diff. %	p-val. t-test	Var.-ratio
Years receiving DI (since 2000)	U	5.3	6.5	-33.4	0.000	0.94	5.4	6.0	-16.2	0.000	0.94
	M	5.3	5.4	-2.1	0.662	0.99	5.4	5.3	4.8	0.429	0.94
Age	U	43.7	46.9	-34.3	0.000	1.09	43.7	46.4	-29.2	0.000	0.94
	M	43.7	43.5	1.2	0.804	1.03	43.7	44.1	-4.3	0.453	1.05
Woman# (ref.: man)	U	47.0%	47.4%	-0.7	0.831	1.00	46.5	47.4%	-1.9	0.659	1.00
	M	47.0%	47.0%	0.0	1.000	1.00	46.5	46.5%	0.0	1.000	1.00
Type of diseases (ref.: congenital defects)											
mental#	U	12.4%	7.3%	17.3	0.000	1.52*	11.1	10.0%	3.6	0.387	1.07
	M	12.4%	11.9%	1.7	0.746	1.02	11.1	9.2%	6.2	0.288	1.13
nervous system #	U	55.7%	54.9%	1.8	0.601	1.00	57.2	53.3%	7.7	0.067	0.99
	M	55.7%	54.7%	2.0	0.674	1.03	57.3	55.7%	3.2	0.587	1.02
musculo-skeletal #	U	4.8%	5.7%	-3.8	0.275	0.85	4.5	5.4%	-4.3	0.326	0.82
	M	4.8%	5.1%	-1.2	0.794	0.95	4.5	4.9%	-1.9	0.743	0.93
other#	U	15.1%	15.8%	-2.0	0.549	0.96	15.0	15.4%	-1.1	0.801	0.98
	M	15.1%	15.4%	-1.0	0.836	0.99	15.1	16.9%	-5.2	0.390	0.91
injuries#	U	6.1%	8.3%	-8.7	0.015	0.75*	5.9	8.0%	-8.3	0.063	0.75*
	M	6.1%	5.7%	1.3	0.756	1.05	5.7	5.2%	1.8	0.731	1.07
Monthly main DI benefits in 1,000 CHF in t-1	U	1217.4	1369.1	-23.7	0.000	1.11	1219.8	1332.0	-17.3	0.000	0.96
	M	1217.4	1228.3	-1.7	0.725	1.01	1221.3	1224.5	-0.5	0.933	1.05
Monthly main DI benefits in 1,000 CHF in t-2	U	1052.8	1283.6	-32.1	0.000	1.12	1060.6	1234.5	-23.9	0.000	0.99
	M	1052.8	1068.0	-2.1	0.666	1.01	1062.4	1033.9	3.9	0.515	1.02
Monthly main DI benefits in 1,000 CHF in t-3	U	954.7	1188.6	-30.6	0.000	1.05	980.9	1141.7	-21.0	0.000	0.96
	M	954.7	979.5	-3.2	0.500	1.01	982.6	955.4	3.5	0.552	0.99
Monthly child DI benefits in 1,000 CHF in t-1	U	208.9	222.9	-2.9	0.396	0.88	228.4	225.2	0.7	0.879	0.90
	M	208.9	236.1	-5.7	0.234	0.87	228.6	270.4	-8.6	0.156	0.89
DI benefit entitlement in %	U	72.6	81.9	-28.8	0.000	1.19	73.0	79.7	-20.3	0.000	0.95
	M	72.6	72.6	0.0	1.000	1.00	73.1	73.0	0.3	0.965	1.00
Civil status: married# (ref.: not married)	U	29.2%	36.4%	-15.5	0.000	0.89	32.5	35.5%	-6.4	0.129	0.96
	M	29.2%	30.6%	-3.0	0.513	1.00	32.4	36.3%	-8.3	0.161	0.97
helplessness allowance # (ref.: no)	U	1.5%	0.8%	7.0	0.012	1.94*	0.9	0.6%	3.1	0.420	1.43*
	M	1.5%	1.7%	-1.4	0.802	0.91	0.9	1.3%	-4.5	0.521	0.70*
supplementary benefit# (ref.: no)	U	34.5%	35.2%	-1.5	0.661	0.99	33.2	33.7%	-1.1	0.790	0.99
	M	34.5%	33.1%	2.9	0.536	1.06	33.2	29.1%	8.6	0.135	1.06
supplementary benefit per case in 1'000 CHF	U	436.2	526.75	-10.5	0.004	0.64*	430.34	495.37	-7.6	0.091	0.75*
	M	436.2	409.04	3.1	0.452	1.02	431.08	410.69	2.4	0.667	0.95
Number of child DI benefits	U	0.43	0.40	2.8	0.399	1.12	0.46	0.41	6.7	0.097	1.19
	M	0.43	0.49	-7.3	0.148	0.86	0.46	0.55	-9.4	0.131	1.03
Extraordinary DI benefits # (ref.: no)	U	5.8%	6.4%	-2.2	0.529	0.92	5.5	6.1%	-2.4	0.574	0.91
	M	5.8%	6.2%	-1.5	0.755	0.96	5.5	5.5%	0.0	0.993	1.00
Nationality# (ref.: Swiss)											
Foreigner: German or Austria#	U	5.1%	3.1%	9.7	0.001	1.57*	4.7	3.1%	7.8	0.040	1.45*
	M	5.1%	5.1%	-0.1	0.979	1.00	4.7	4.2%	2.6	0.679	1.11

45

Variable											
Foreigner: EU or EFTA countries#	U	13.7%	13.7%	-0.2	0.945	1.00	14.2	13.7%	1.2	0.768	1.03
	M	13.7%	13.8%	-0.3	0.943	0.99	14.2	12.5%	4.9	0.397	1.15
Foreigner: rest of the world#	U	23.0%	25.4%	-5.5	0.107	0.93	25.2	25.4%	-0.4	0.932	1.00
	M	23.0%	22.5%	1.1	0.806	1.02	25.1	26.3%	-2.7	0.647	0.97
Daily allowance of the DI in t-1# (ref.: no)	U	3.6%	1.2%	16.0	0.000	2.90**	3.6	1.6%	12.4	0.000	2.09**
	M	3.6%	4.7%	-7.2	0.240	0.83	3.6	1.8%	11.6	0.051	1.56*
Daily allowance of the unemployment insurance in t-1# (ref.: no)	U	13.1%	2.3%	41.1	0.000	3.74**	10.5	2.6%	32.6	0.000	2.93**
	M	13.1%	11.3%	6.7	0.252	1.03	10.4	10.4%	0.1	0.987	1.00
Monthly income from paid employment in 1,000 CHF in t-1	U	748.8	757.4	-0.6	0.868	0.64*	766.6	801.2	-2.4	0.608	0.60*
	M	748.8	859.5	-7.9	0.064	0.86	765.8	775.9	-0.7	0.886	1.22
Monthly income from paid employment in 1,000 CHF in t-2	U	917.2	865.3	3.3	0.364	0.71*	828.1	878.6	-3.2	0.486	0.66*
	M	917.2	999.5	-5.2	0.262	0.77*	827.4	871.8	-2.8	0.606	0.87
Monthly income from paid employment in 1,000 CHF in t-3	U	1212.5	916.8	15.6	0.000	0.92	1089.6	993.3	5.1	0.261	0.69*
	M	1212.5	1295.4	-4.4	0.382	0.76*	1084.7	1151.3	-3.5	0.509	1.04
Income from paid employment in t-1# (ref.: no)	U	68.3%	55.7%	26.2	0.000	0.93	75.1	62.1%	28.4	0.000	0.87
	M	68.3%	70.3%	-4.3	0.343	1.06	75.1	72.8%	5.1	0.367	0.94
Income from paid employment in t-2# (ref.: no)	U	67.4%	66.8%	1.2	0.724	0.98	59.6	61.7%	-4.3	0.302	1.02
	M	67.4%	68.7%	-2.9	0.539	1.04	59.5	58.6%	1.8	0.764	1.00
Income from paid employment in t-3# (ref.: no)	U	72.8%	57.7%	32.2	0.000	0.85	76.9	63.0%	30.6	0.000	0.83
	M	72.8%	75.9%	-6.6	0.130	1.14	76.8	74.8%	4.6	0.413	0.90
Participation in VR measures in t-1# (ref.: no)	U	16.0%	2.0%	50.6	0.000	4.66**	6.4	1.3%	26.8	0.000	4.01**
	M	16.0%	16.9%	-3.5	0.579	0.97	6.2	5.7%	3.0	0.680	1.07
Participation in VR measures in t-2# (ref.: no)	U	4.4%	1.5%	17.3	0.000	2.81**	4.0	1.6%	14.3	0.000	2.34**
	M	4.4%	4.7%	-1.7	0.768	0.95	3.8	3.6%	1.6	0.818	1.06
Participation in VR measures in t-3# (ref.: no)	U	5.3%	1.7%	19.4	0.000	2.87**	5.4	1.9%	18.3	0.000	2.57**
	M	5.3%	5.4%	-0.7	0.907	0.98	5.2	5.6%	-2.1	0.764	0.93
Functional disorder# (ref.: others)											
impairment of general condition#	U	40.6%	40.9%	-0.4	0.898	1.00	40.8	39.6%	2.5	0.557	1.01
	M	40.6%	38.7%	4.0	0.387	1.03	40.7	41.7%	-2.1	0.724	1.00
behavioral disorders#	U	21.4%	22.0%	-1.6	0.633	0.97	23.7	21.6%	5.0	0.224	1.08
	M	21.4%	21.5%	-0.3	0.947	0.99	23.7	22.1%	3.7	0.530	1.08
multiple mental disorders#	U	8.0%	7.8%	1.0	0.774	1.03	7.9	7.5%	1.8	0.669	1.06
	M	8.0%	8.0%	0.2	0.972	1.01	8.0	8.7%	-2.8	0.648	0.93
at the trunc#	U	4.4%	4.8%	-1.8	0.600	0.92	4.3	4.7%	-1.8	0.672	0.92
	M	4.4%	5.7%	-6.3	0.197	0.80	4.3	5.3%	-4.5	0.453	0.83
multiple mental and physical disorders#	U	4.2%	5.0%	-3.8	0.277	0.85	4.3	4.9%	-2.6	0.548	0.89
	M	4.2%	3.9%	1.2	0.794	1.06	4.3	4.8%	-2.1	0.718	0.91
Year dummies# (ref.: 2009)											
2010#	U	58.0%	33.3%	51.2	0.000	1.62*	91.0	50.0%	100.7	0.000	0.77*
	M	58.0%	57.9%	0.3	0.957	1.00	91.0	91.0%	0.0	1.000	1.00
2011#	U	36.2%	33.3%	6.1	0.067	1.05					
	M	36.2%	36.3%	-0.1	0.978	1.00					

Notes: See Table 6.