

Hagen, Tobias

**Conference Paper**

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

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Health I, No. B09-V3

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Hagen, Tobias (2016) : Econometric Evaluation of a Placement Coaching Program for Recipients of Disability Insurance Benefits in Switzerland, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Health I, No. B09-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/145736>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

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

## Abstract

*This paper evaluates a placement coaching program carried out in Zurich during 2009–2013 that focused on the reemployment of persons drawing disability insurance (DI) benefits. A private company was commissioned with carrying out the program. Kernel-based matching and radius matching with bias adjustment (Lechner et al. 2011; Huber et al. 2015) estimators combined with difference-in-differences are applied to administrative panel data. The estimates indicate 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 government. Sensitivity analyses indicate that the results are robust to confounders. An interesting policy implication is that it seems possible to enhance the employment prospects of disabled persons with a relatively inexpensive placement coaching measure which does not include any explicit investments in human capital.*

**JEL Classification:** I38; J08; J14, J64

**Keywords:** Rehabilitation; Placement Coaching; Disability; Evaluation; Matching

## I. Introduction

Many countries face the problem of high costs for disability insurance (DI) benefits due to an increased number of recipients. This led to reforms aiming to increase the employment rate of (potential) recipients (Burkhauser et al. 2014, OECD 2010, Bound and Burkhauser 1999). Basically there are two ways to achieve this: either the inflows onto DI benefits can be reduced (which seems to be the major attempt) or the outflows from DI benefits into employment (Burkhauser et al. 2014). As mentioned 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 from the DI system, including the possible loss of the 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 their learned profession, or, if this is not possible any more, to train these persons to give them the qualification for new jobs (OECD 2010).

In contrast to these VR measures, a pilot project was carried out in Switzerland (canton Zurich) during 2009–2013, which did not include any explicit investment in human capital or medical aid, but placement coaching by a private company being specialized in this field. As a result the pilot project was relatively inexpensive compared to training measures. In this paper a microeconomic 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 available covering the employment history of the persons. The application of DiD and the conditioning on pre-treatment outcome variables seems valid as anticipatory effects are unlikely.

Empirical evidence on the effectiveness of job placement services for recipients of DI benefits is scarce. Firstly, there are several studies on the effectiveness of VR measures for recipients of DI benefits. Secondly, there are a lot of evaluation studies for 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 compared to non-participation of measure 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 for men. This study also attempt to provide insights into the costs and benefits of the program. From the perspective of the program, the expected benefits (reductions in payments) exceed the costs for women. Aakvik et al. (2005) evaluate VR programs on employment for female applicants in Norway. They find a negative effect of the training program. 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 program administrators only infrequently select individuals with unfavorable employment prospects into the measure. For the 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.<sup>1</sup> They find large positive long-run (3 to 9 years) effects on employment and earnings. Both the short- and long-run mean labor market effects are estimated to be positive for diagnosis & evaluation, training, education and other services, but negative for restoration and maintenance. The mean long-run benefits exceed the mean costs by 4 to 6 times. For Norway Markussen and Røed (2014) evaluate four different VR programs for temporary disability insurance 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

<sup>1</sup> An explanation of the services is given in Dean et al. (2015) on Page 403:

*“Diagnosis & evaluation* are provided at intake in assessing eligibility and developing an IPE [individualized plan for employment]. *Training* includes vocationally oriented expenditures for on the job training, job coach training, work adjustment, and supported employment. *Education* includes tuition and fees for a GED (graduate equivalency degree) program, a vocational or business school, a community college, or a university. *Restoration* covers a wide variety of medical expenditures including dental services, hearing/speech services, eyeglasses and contact lenses, drug and alcohol treatments, psychological services, surgical procedures, hospitalization, prosthetic devices, and other assistive devices. *Maintenance* includes cash payments to facilitate everyday living and covers such items as transportation, clothing, motor vehicle and/or home modifications, and services to family members. *Other services* consists of payments outside of the previous categories such as for tools and equipment.”

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, it can be said that, these kind of measures seem to be successful in terms of helping the participants to increase their employment prospects and in terms of a reducing government spending.

There is a lot of empirical evidence on 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 the counselling, but also because they may be associated with “threat effects” for beneficiaries who risk sanctions in case of lack of search effort. Threat effects did not exist at all in the pilot project in Zurich since sanctions were not a part of 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 for 9 European countries and find that job search assistance programs decrease unemployment duration and increase the employment rates. These statements are 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 a certain similarity to the pilot project in Zurich. Høgelund and Holm (2006) evaluate the effect of case management interviews (CMI) performed by social caseworkers on the probability of return to work of disabled employees. Based on instrumental variables and a competing hazard rate model, they find that CMI increases the probability of returning to work for the pre-sick leave employer, but has no effect on the probability of resuming work for a new employer. However, as the CMI is made by case managers of the municipalities they have all the instruments of VR available which 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, 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 the 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) exceed these costs.

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 into it. The former is important for the understanding of possible causal mechanisms on 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 the application to the subject of 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 government’s perspective. The paper concludes with a summary and the discussion of policy implications.

## **II. Institutional background: the Swiss Disability Insurance**

The Swiss social security system in general and 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 insurances, which are mandatory for (almost) all employees. The third pillar is the private provision of employees, which should complement the first two pillars. The third pillar is also protected by law and often promoted by tax facilities.

The pilot project and all administrative data are related to the first pillar. This implies that information on the second and the third pillar are not available. This is an obvious methodological shortcoming of this paper because confounding variables from the second and the third pillar may affect incentives and, hence, the selection process 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 disability insurance (DI), as a part of the first pillar, aims at guaranteeing the existence needs of insured persons who have become disabled by paying DI benefits and/or by providing

rehabilitation measures.<sup>2</sup> Disability is defined as a decline in the ability to earn a living or in the accomplishment of daily tasks such as housework resulting in physical, psychological or mental health problems. This incapacity must last at least one year. For the judgement of the presence of an inability to earn a living it does not matter what are the causes of the health depreciation. 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 contributed for at least three whole years can lay claim to a DI benefit. The right to a DI benefit begins when the insured person has an average incapacity to work of at least 40% (so-called *disability degree*) and after a waiting period of one year following the end of the 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, ranging from 25% to 100% (Table 1).

Moreover, Table 1 provides a statistic on the estimation sample. It can be seen that the participants are a positive selection with regard to the degree of disability: The mean DI benefit entitlement is 72.6% versus 81.9% in case of the non-participants. Nevertheless, one should keep in mind that more than half 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 %  |  |
|-------------------------------------|------------------------|---|--|
| Degree of disability                | DI benefit entitlement | Non-participants<br>(potential control<br>observations) | Participants in the year<br>before the individual<br>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.                      |                        | 40710   | 908  |

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

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

### **III. The Pilot Project and the Selection Mechanism**

For the pilot project an international private company focusing on workforce participation was commissioned. The measure 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 and practical tips in 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 only for 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 to provide financial incentives or to sanction their clients' 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” means here that participants did not show up at appointments with their advisers or that they indicated that they did not want to participate anymore. Hence, part of the dropouts may be persons who founds successfully jobs on their own, but 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”). Those 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). 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 this is the legal regulation 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 8,740.

In summary, the pilot project carried out by a privat company 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 possibilities to provide additional financial incentives or make use of “threat effects”. (3) Due to the bonuses, the company had financial incentives not only to bring the participants in (higher paid) jobs, but also to keep them employed for 52 weeks.

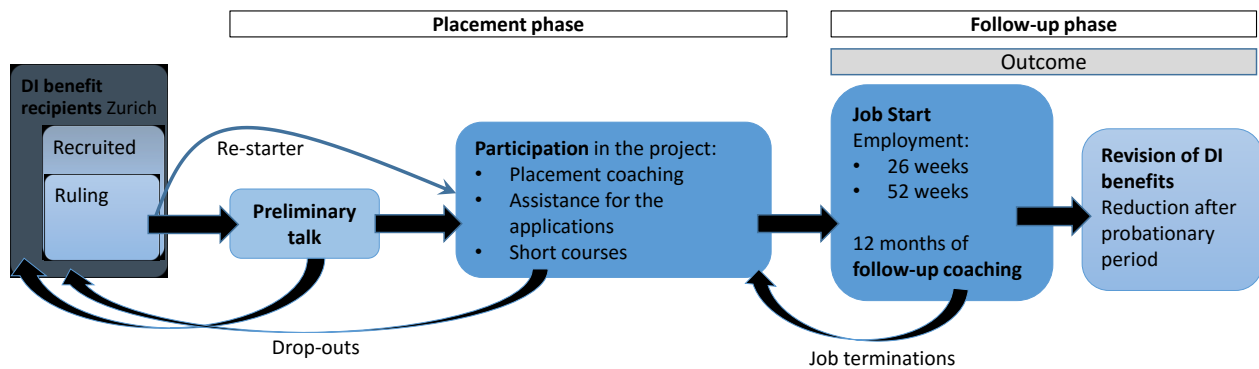


FIGURE 1: FLOWCHART OF THE PILOT PROJECT

Before DI benefit recipients could take part in the pilot project, they had to complete the following three-stage process (on the far-left in Figure 1):

In the first stage, the DI office in Zurich recruited potential participants from the population of almost 50,000 DI benefit recipients in Zurich, that means, they were informed on the possibility to participate and the implications for their entitlement. People aged between 18 and 58 were targeted, the intention being to achieve an age distribution reflecting that of the entire population of DI benefit recipients. The participants needed to exhibit reintegration potentials. 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 as well as those who volunteered for the pilot projects and signed a corresponding agreement with the Zurich DI office. Insured persons presumed to be incapable of 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 group of the participants as well as the control group generated by the matching estimators. Put differently, the almost 35,000 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 with a participation permission.

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 a participation. A total of 947 persons took part in the project between November 2009 and August 2011. For reasons of methodology 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 time 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.), the participation in rehabilitation measures and income. The variables are discussed in greater detail below.

The number of persons is shown in Table 2 for December 2009. The total number of recipients in the sample at the end of the year 2009 was almost 50,000. Out of these 14,878 persons 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 persons died until 2014 are already excluded. This represents 525 (3.4%) of all the recruited persons. For the sub-group of participants this amounts to 19 persons (2%). Furthermore, persons who transited into the old-age pension system until 2014 are excluded. This is the case for 24 (0.16%) recruited persons and for 8 (0.87%) participants.

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

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

Notes: \* Persons died until Dec 2014 and people transited into the old-age pension 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 transited to the old-age pensions 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<br>(program starters) | Nonparticipants<br>(potential controls) | Total          |
|-------|------------------------------------|---|----------------|
| 2009  | 52<br>0.38%                        | 13,570<br>99.62%                        | 13,622<br>100% |
| 2010  | 527<br>3.74%                       | 13,570<br>96.26%                        | 14,097<br>100% |
| 2011  | 329<br>2.37%                       | 13,570<br>97.63%                        | 13,899<br>100% |
| Total | 908<br>2.18%                       | 40,710<br>97.82%                        | 41,618<br>100% |

As explained in greater detail below, the applied approach strives for an annual accuracy only. The potential control group is based on persons who were recruited but did not participate (14,878 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 potential controls being available each year.

TABLE 4—OUTCOME VARIABLES

| Outcome Variable  | Period                   | Adjusted for   | Notes  |
|---|--------------------------|--|--|
| Monthly main DI benefit in CHF  | Dec. 2000<br>– Dec. 2014 | DI benefit increases<br>(take place every second year) | measured in Dec. each year                                 |
| Monthly total DI benefit in CHF<br>(main benefit and child's benefit) | Dec. 2000<br>– Dec. 2014 | DI benefit increases<br>(take place every second year) | measured in Dec. each year                                 |
| DI benefit entitlement in %   | Dec. 2000<br>– Dec. 2014 |  | measured in Dec. each year; see Table 1 for an explanation |
| Monthly supplementary benefits<br>(SB) per case in CHF                | Dec. 2000<br>– Dec. 2014 | Consumer price index, annual                           | total payment for households                               |
| DI benefit recipient (yes) in %                                       | Dec. 2000<br>– Dec. 2014 |  | Dummy variable   |
| Annual income earned from paid<br>employment in CHF                   | 2000 – 2013              | Consumer price index, annual                           | Calculated from contribution to social security            |
| Income earned from paid<br>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. In contrast, the income variables are measured with a daily frequency, but they are aggregated to annual data. 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 in the administrative panel dataset used (see Table 7 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 before reaching the age of 18, or until the completion of the education with an maximum age of 24. Variables on the educational background are not available.
- **Health status:** There are no subjective information available with regard to well-being. However, all the information which are necessary for the application and the approval of a DI benefit are 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*, *at the trunc*, *multiple mental and physical disorders*) and the others are pooled into one residual category. Furthermore, the data include the information on the so-called *helplessness allowance*, for which people who permanently require a considerable degree of help from a third person are entitled to. Interestingly, 1.5% of the participants receipted helplessness allowance in the year before their individual start of the project.
- **Occupational history:** Total *income* subject to deduction of social insurance contributions and *income from paid employment* are available since the year 2000. Furthermore, previous *participations in VR* measures of the DI system can be identified. The *daily allowance* of the DI system also indicate that a person participated in a VR measure. The receipt of a *daily*

*allowance of the unemployment insurance* indicate that a person has paid enough contribution and is still in the labor market.

- **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 having become disabled before their 20<sup>th</sup> 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). The next Section will discuss how to deal methodologically with this problem of missing variables and the possibility of selection on unobservables. Due to its long time dimension, the dataset seems nevertheless to be relatively comprehensive. This becomes obvious when comparing it, e.g., to 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 drop-outs from the program), in comparison to non-participation in the program, which, however, have been recruited before by the DI office (see Table 2). The non-participants were selected by the DI office as potential participants, but decided not to participate. They had the possibility to participate in another public-provided program. The latter is a reasonable definition since the evaluated measure 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 has been criticized for leading to biased estimates: If, based on the 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. The persons could start the program between December 2009 and August 2011. *Secondly*, there was no excess demand and hence no queue of persons waiting on their opportunity to participate. In contrast, there was the tendency that it was hard to fill all available positions: out of 14,878 informed 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) which 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, non-participants can be defined as persons who never participate in that program, but possibly 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  $X$ .<sup>3</sup> 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  $\hat{e}_1(X)$  and  $j = 1, \dots, N_0$  and untreated individuals  $\hat{e}_0(X)$  is predicted. The common support condition is fulfilled, if for every treated  $i$  the  $\hat{e}_1(X)$  is overlapped by  $\hat{e}_0(X)$ .

After matching the ATT effect is calculated as  $ATT = \frac{1}{N_1} \sum_{i=1}^{N_1} (Y_{1i} - \sum_{j=1}^{N_0} w(i, j) Y_{0j})$ , 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  $w(i, j) = \frac{G\left(\frac{\hat{e}_j - \hat{e}_i}{h}\right)}{\sum_{k \in (C=0)} G\left(\frac{\hat{e}_k - \hat{e}_i}{h}\right)}$ , where  $G(\cdot)$  is the kernel function and  $h$  is the bandwidth parameter. Since it leads to the best results 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

<sup>3</sup> 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.

Epanechnikov kernel is applied here. This kernel has the advantage to put a weight of zero to control observations outside the bandwidth (see Galdo et al. 2008).

There is a growing 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.<sup>4</sup> Surprisingly and in contrast to the literature, the estimated ATT effects here are relatively robust to changes in the bandwidth (see Section VII).

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 an 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. By doing this, it is guaranteed that controls must have the same gender and the same DI benefit entitlement in % as their corresponding treated individuals. Furthermore, control observation come from the same calendar year as program starts of the 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 infeasible 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  $ATT = \frac{1}{N_1} \sum_{i=1}^{N_1} \left( (Y_{1i\tau} - Y_{1i,t-1}) - \sum_{j=1}^{N_0} w(i,j) (Y_{0j\tau} - Y_{0j,t-1}) \right)$  with  $\tau = \{t + 1, t + 2, t + 3, t + 4\}$  and  $t$  being the year of the individual program start (2009, 2010 or 2011). In contrast to the calculation of the ATT effect as difference in post-program “levels” of outcome variables ( $Y_1 - Y_0$ ), the ATT effect is now calculated as the differences of changes of  $Y_1$  and  $Y_0$  over time. For the example of the ATT effect in  $t + 1$ , this means for the treatment group

<sup>4</sup> This method of bandwidth choice is applied inter alia by Heckman et al. (1998b).

$(Y_{1,t+1} - Y_{1,t-1})$  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 the 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 “pillar” 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 especially risk attitude may be relevant here: if a DI benefit recipient takes up a sufficiently well-paid job he or she may lose he or her entitlement to DI completely. This implies that this person must go through the whole application process again if he or she becomes disabled again in the future.

However, DiD and conditioning on pre-treatment outcome variables in general, requires that there are no “anticipatory effects”. For example, an anticipatory effect may mean that future participants in retraining measure for unemployed workers reduce their search effort already 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.

### ***B. Application of Propensity Score Matching to the Evaluation Problem Here***

Due to the data availability the whole analyses are based on two samples. The reason is that the outcome variables for DI benefit are available until December of 2014; the outcome variables for incomes are available until end of 2013. The individual program start 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 starting in 2009, 2010 and 2011. This sample includes 908 persons.

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 starting in 2009 and 2010. This sample includes 579 persons (=52 starters from 2009 + 527 starters from 2010).

As mentioned in the last 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 is based on the following considerations. On the one hand, it seems plausible to include all available variables including the pre-treatment outcome variables into the conditioning set since all variables are potential determinants for outcome as well as selection into the 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 suggests that also 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 the discussion in Caliendo and Kopeinig 2008, Millimet and Tchernis 2009, Austin 2011).

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 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 which 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 statistical significant effect in the probit. The results of the propensity score estimate will be presented in the next section.

As explained in the last Section V.A., the inclusion of lagged outcome variables and the DiD estimator require 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. Neither corresponding eligibility criteria were in force, nor was there any direct financial advantage or disadvantage 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  $Y_{t-1}$  and  $Y_{t-2}$  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:

(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. Generally, the labor market conditions in Zurich are excellent from an international perspective. From 2009 to 2014 there were some variations in the overall unemployment rate in Zurich.<sup>5</sup> However, it is unclear whether and to what extent this is relevant for the DI benefit recipients. In case of the method radius matching with bias adjustment, the starting year  $t$  is included into the Mahalanobis distance (besides 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 the number of potential controls to 13,570 per year (Table 3).

(2) *Ignore  $t$  and allow that all 40,710 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 into the explanatory variables of the propensity score.* Hence,  $t$  is one conditioning variable besides 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-program outcome variables. However, it comes with the cost, that  $t$  is not identical for all treated and corresponding controls.

After the estimation and prediction of the propensity score the Common Support is examined. For every participant  $i$  it is checked whether  $\hat{e}_1(X)$  is overlapped by  $\hat{e}_0(X)$ . In the “program starters 2009–2011”-sample this given for all  $N_1=908$  participants. In the “program starters 2009–2010”-sample this is true for all but one persons and thus the estimates are based on 578 treated persons only. 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.

<sup>5</sup> 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](http://www.amstat.ch).

## VI. Empirical Results

### A. Propensity Score

Table 5 shows the estimation results of the propensity score probit for both samples. Due to high collinearity many coefficients cannot be interpreted. Nevertheless, there are some interesting results which should be mentioned. There are no gender-differences in the participation probability. 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

| Conditioning variables<br>Lagged values (previous calendar year), unless specified differently. | Coefficient<br>(t-stat.)         |                                  |
|---|----------------------------------|----------------------------------|
|   | Program<br>starters<br>2009-2011 | Program<br>starters<br>2009-2010 |
| Years in DI benefit (since 2000)  | -0.0278***<br>(-4.36)            | -0.0131<br>(-1.58)               |
| Age   | 0.0497***<br>(3.60)              | 0.0743***<br>(4.19)              |
| Age squared / 1,000   | -0.736***<br>(-4.57)             | -0.0010***<br>(-4.93)            |
| Woman <sup>#</sup> (ref.: man)  | -0.0047<br>(-0.15)               | -0.0241<br>(-0.63)               |
| Type of diseases (ref.: congenital defects)   |                                  |                                  |
| mental <sup>#</sup>   | -0.101<br>(-1.17)                | -0.0645<br>(-0.53)               |
| nervous system #  | -0.140<br>(-1.37)                | 0.0303<br>(0.36)                 |
| musculo-skeletal #  | -0.0966<br>(-1.10)               | 0.00238<br>(0.02)                |
| other <sup>#</sup>  | -0.196**<br>(-2.04)              | 0.0501<br>(0.56)                 |
| injuries <sup>#</sup>   | -0.127<br>(-1.32)                | -0.0653<br>(-0.62)               |
| Monthly main DI benefits in 1,000 CHF in $t - 1$  | 0.0810<br>(1.24)                 | 0.0674<br>(0.82)                 |
| Monthly main DI benefits in 1,000 CHF in $t - 2$  | -0.125***<br>(-2.60)             | -0.138**<br>(-2.23)              |
| Monthly main DI benefits in 1,000 CHF in $t - 3$  | 0.0569<br>(1.34)                 | 0.0628<br>(1.14)                 |
| Monthly child DI benefits in 1,000 CHF in $t - 1$   | -0.110<br>(-1.47)                | -0.0698<br>(-0.78)               |
| DI benefit entitlement in %   | -0.0015<br>(-1.31)               | -0.0033**<br>(-2.42)             |
| Civil status: married <sup>#</sup> (ref.: not married)  | -0.0591<br>(-1.41)               | -0.0202<br>(-0.39)               |
| Helplessness allowance # (ref.: no)   | 0.367***<br>(2.69)               | 0.289<br>(1.34)                  |
| Supplementary benefit <sup>#</sup> (ref.: no)   | 0.146***<br>(2.79)               | 0.111*<br>(1.67)                 |
| Supplementary benefit per case, amount in 1'000 CHF   | -0.109***                        | -0.093***                        |

|  |                       |                      |
|--|-----------------------|----------------------|
|  | (-3.75)               | (-2.58)              |
| Number of child DI benefits  | 0.0826*<br>(1.94)     | 0.0606<br>(1.20)     |
| Extraordinary DI benefit <sup>#</sup> (ref.: no)                         | -0.140*<br>(-1.94)    | -0.0729<br>(-0.81)   |
| Nationality <sup>#</sup> (ref.: Swiss)                                   |                       |                      |
| Foreigner: German or Austria <sup>#</sup>                                | 0.161**<br>(2.13)     | 0.140<br>(1.46)      |
| Foreigner: EU or EFTA countries <sup>#</sup>                             | -0.0270<br>(-0.57)    | -0.0114<br>(-0.20)   |
| Foreigner: rest of the world <sup>#</sup>                                | -0.0582<br>(-1.38)    | -0.0192<br>(-0.37)   |
| Daily allowance of the DI in $t - 1^{\#}$ (ref.: no)                     | -0.0540<br>(-0.49)    | 0.0155<br>(0.11)     |
| Daily allowance of the unemployment insurance in $t - 1^{\#}$ (ref.: no) | 0.577***<br>(9.29)    | 0.495***<br>(6.13)   |
| Monthly income from paid employment in 1,000 CHF in $t - 1$              | -0.065***<br>(-3.21)  | -0.046*<br>(-1.82)   |
| Monthly income from paid employment in 1,000 CHF in $t - 2$              | -0.017<br>(-0.79)     | -0.021<br>(-0.86)    |
| Monthly income from paid employment in 1,000 CHF in $t - 3$              | 0.0131<br>(1.42)      | -0.00280<br>(-0.17)  |
| Income from paid employment in $t - 1^{\#}$ (ref.: no)                   | 0.0729*<br>(1.67)     | 0.00686<br>(0.12)    |
| Income from paid employment in $t - 2^{\#}$ (ref.: no)                   | 0.100***<br>(2.61)    | 0.167***<br>(3.58)   |
| Income from paid employment in $t - 3^{\#}$ (ref.: no)                   | 0.0589<br>(1.43)      | 0.0241<br>(0.44)     |
| Participation in VR measures in $t - 1^{\#}$ (ref.: no)                  | 0.909***<br>(5.89)    | 0.540***<br>(5.27)   |
| Participation in VR measures in $t - 2^{\#}$ (ref.: no)                  | 0.218**<br>(2.38)     | 0.225*<br>(1.92)     |
| Participation in VR measures in $t - 3^{\#}$ (ref.: no)                  | 0.284***<br>(3.37)    | 0.294***<br>(2.83)   |
| Functional disorder <sup>#</sup> (ref.: others)                          |                       |                      |
| Impairment of general condition <sup>#</sup>                             | 0.0721<br>(1.38)      | 0.146**<br>(2.19)    |
| Behavioral disorders <sup>#</sup>  | 0.0842<br>(1.33)      | 0.203**<br>(2.53)    |
| Multiple mental disorders <sup>#</sup>                                   | 0.0403<br>(0.53)      | 0.135<br>(1.40)      |
| At the trunc <sup>#</sup>  | 0.160*<br>(1.82)      | 0.156<br>(1.41)      |
| Multiple mental and physical disorders <sup>#</sup>                      | 0.0652<br>(0.76)      | 0.140<br>(1.31)      |
| Year dummies <sup>#</sup> (ref.: 2009)                                   |                       |                      |
| 2010 <sup>#</sup>  | 0.984***<br>(17.71)   | 0.981***<br>(17.32)  |
| 2011 <sup>#</sup>  | 0.755***<br>(13.30)   |                      |
| Constant   | -3.281***<br>(-11.11) | -3.819***<br>(-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 in the sample                 | 2.178%                | 2.087%               |
| Average predicted participation probability (Propensity Score)           | 2.182%                | 2.089%               |

Notes: <sup>#</sup>Dummy-Variable.  $t$  statistics in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## ***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 the “one” balancing test, but different approaches have been proposed. Lee (2013) provides an overview. Some of the procedures implemented by Leuven and Sianesi (2003) in the STATA command “pstest” are applied here.

In Table 7 shows detailed analysis for every covariate included in the propensity score. To keep it simple the detailed results are presented for kernel-based matching only. Table 7 shows for both time periods (program starters 2009–2011 vs. 2009–2010) as well as the unmatched (U) and the matched (M) samples balancing tests with respect to each covariate. First of all, the mean of each variable in the treated group ( $\bar{x}_1$ ) and the untreated group ( $\bar{x}_0$ ) 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 7 ) 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) argues 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 7 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 “balanced” this variance-ratio is near 1.0. Again, it can be seen that the measure is close to 1.0 for all variables in the matched samples. Moreover, the difference to 1.0 is smaller than in the unmatched samples for almost all variables.

The next step is to look at summary measures of the overall (im)balance of all conditioning variables. These are shown in Table 6 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 have no predictive power for the participation anymore, that means, differences between treated and untreated individuals are balanced. The p-value of the Likelihood Ratio test of the joint significance of all explanatory variables in the probit model points into the same direction. The following two columns in Table 6 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 the both final columns of Table 6 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, what is again given 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, also 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 with bias adjustment estimator in terms of balancing the conditioning variables. This is true for all summary measures in Table 6 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, SUMMARY MEASURES

|                 |   | Program starters 2009-2011 |                                  |           |           |           |           |
|-----------------|---|----------------------------|----------------------------------|-----------|-----------|-----------|-----------|
| Matching Method |   | Pseudo-R <sup>2</sup>      | p-value<br>Likelihood Ratio Test | Mean Bias | Med. Bias | Rubin's B | Rubin's R |
| -               | 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      |

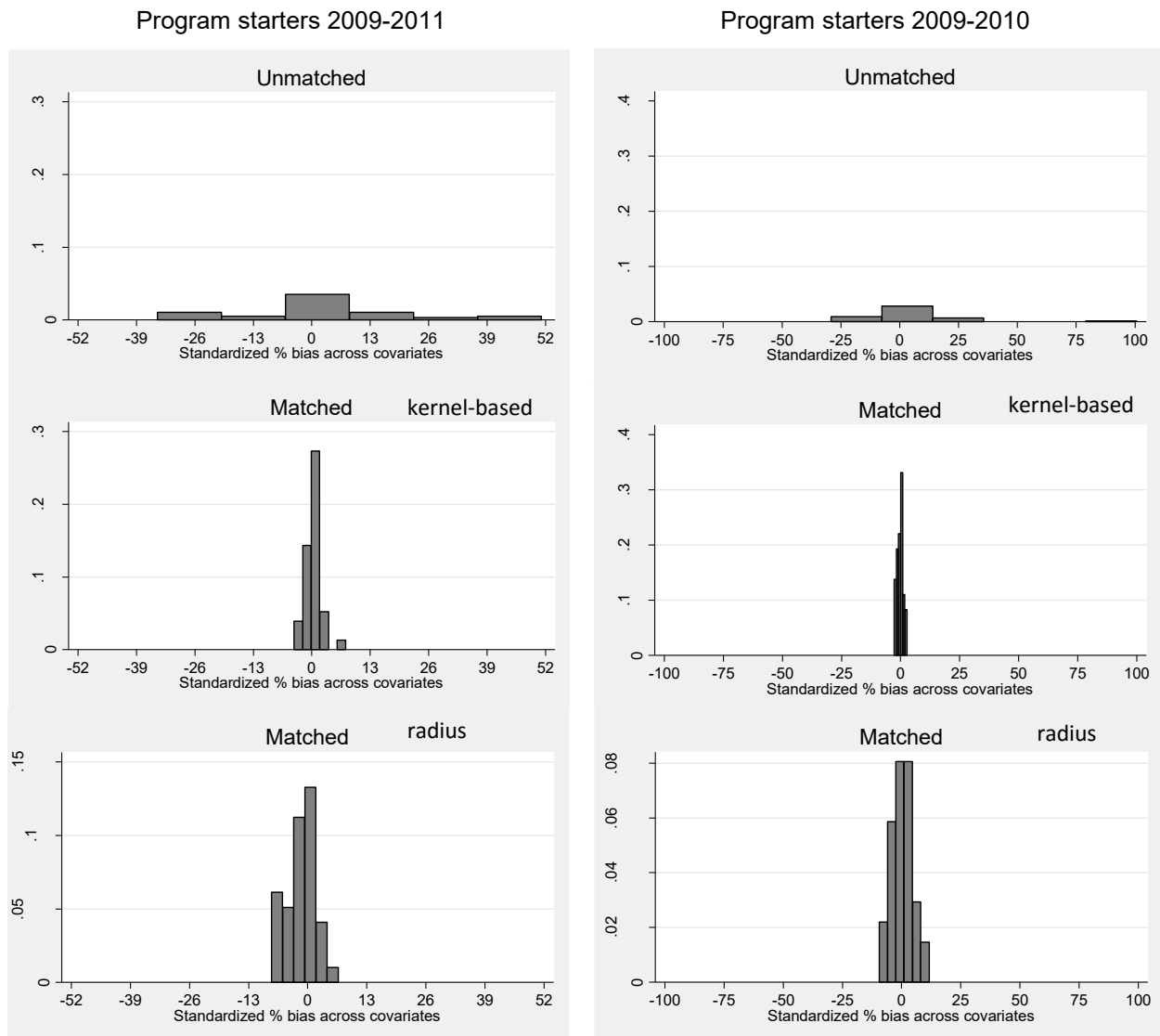


FIGURE 2: HISTOGRAM OF STANDARDIZED DIFFERENCES OF ALL CONDITIONING VARIABLES IN TABLE 7

TABLE 7 —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 in 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 <sup>#</sup> (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 <sup>#</sup>                                    | 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 <sup>#</sup>                                     | 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 <sup>#</sup>                                  | 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 <sup>#</sup> (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 <sup>#</sup> (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 <sup>#</sup> (ref.: Swiss)                 |   |                            |             |              |               |            |                            |             |              |               |            |
| Foreigner: German or Austria <sup>#</sup>              | 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 <sup>#</sup>                                 | 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 <sup>#</sup>                                    | 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 <sup>#</sup> (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 <sup>#</sup> (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 <sup>#</sup> (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 <sup>#</sup> (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 <sup>#</sup> (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 <sup>#</sup> (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 <sup>#</sup> (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 <sup>#</sup> (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 <sup>#</sup> (ref.: others)                              |   |        |        |      |       |        |        |        |       |       |        |
| impairment of general condition <sup>#</sup>                                 | 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 <sup>#</sup>  | 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 <sup>#</sup>                                       | 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 <sup>#</sup>  | 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 <sup>#</sup>                          | 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 <sup>#</sup> (ref.: 2009)                                       |   |        |        |      |       |        |        |        |       |       |        |
| 2010 <sup>#</sup>  | 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 <sup>#</sup>  | U | 36.2   | 33.3   | 6.1  | 0.067 | 1.05   |        |        |       |       |        |
|  | M | 36.2   | 32.6   | 7.6  | 0.106 | 0.98   |        |        |       |       |        |

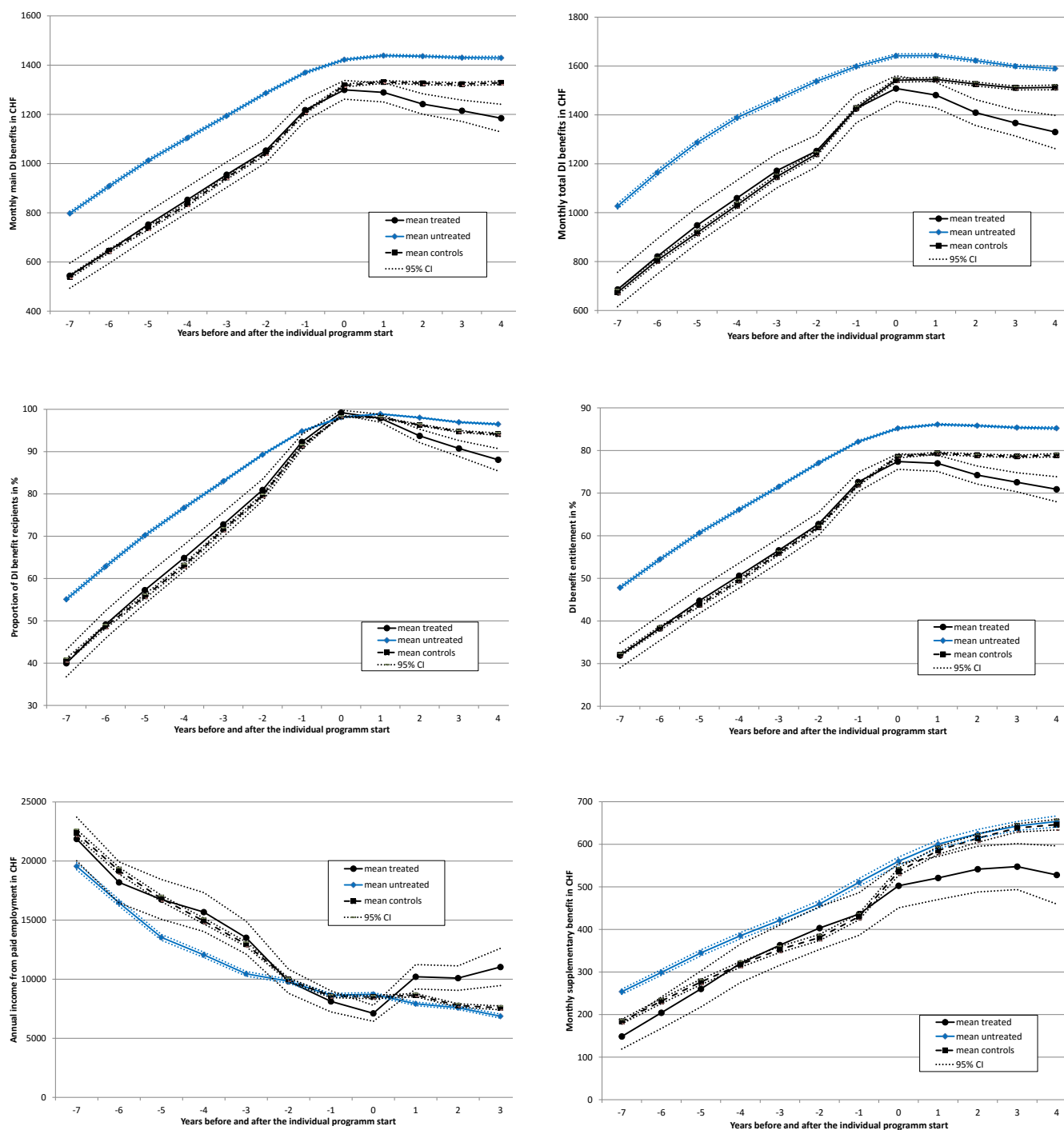
Notes: <sup>#</sup>Dummy-Variable. Standardized difference  $\frac{|\bar{x}_1 - \bar{x}_0|}{\sqrt{(V(x_1) + V(x_0))/2}}$  (Rosenbaum and Rubin, 1985). Ratio of the variance of the residuals orthogonal to the linear index of the propensity score in the treated group over the non-treated group (Rubin, 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.

## 2. Pre-Program Outcome Variable

The “Pre-Program test” proposed by Heckman and Hotz (1989) is based on the consideration that, if there is no selection bias left, there should be no significant differences between the mean outcome variables of treated and the controls before the individual start of the participation. Put differently, the ATT before the treatment should be zero. However, this test is unsuitable, if anticipatory effects are relevant. As explained in Section V.B and as can be seen in the following figures anticipatory effects are unlikely.

The following analyses are shown for the kernel-based matching estimator only, which is defined here as the “preferred specification” (see last 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 of the program  $t = \{2009, 2010, 2011\}$ . The graph 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 graph middle left). 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) there are no statistically significant differences between treated and controls left. Hence, with 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 graph top right in Figure 3. Furthermore, both graphs on the top 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 graph bottom left shows that 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, that 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 difference-in-differences (“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 result of the three methods differ only slightly which may be an indication for the robustness. Focusing on the “Kernel, DiD”-estimates and interpreting the results one may draw the following conclusions: All the findings indicate that, on average, the measure is effective for the participants in that the DI amounts paid to them is lowered by raising their levels of paid employment. Compared with the control group, the proportion of DI recipients among the participants was 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%.<sup>6</sup> 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  $t + 4$ . That 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 lead 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  $t + 4$ .

<sup>6</sup> This relative increase of 37% can be roughly calculated as follows: From the graph at the bottom-left of Figure 3 it can be seen that the counterfactual income is approximately CHF 7,500. Then CHF 2,750 / CHF 7,500 is approximately 37%.

TABLE 8: ESTIMATED ATT-EFFECTS ON **DI BENEFITS AND SB** – KERNEL-BASED MATCHING, T-STAT IN PARENTHESES

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

TABLE 9: ESTIMATED ATT-EFFECTS ON **INCOME** – T-STAT IN PARENTHESES

| Outcome Variable                                 | Method         | $t + 1$            | $t + 2$            | $t + 3$            |
|--|----------------|--------------------|--------------------|--------------------|
|  |                | starters           | starters           | starters           |
|  |                | 2009-2011          | 2009-2011          | 2009-2010          |
|  |                | outcome            | outcome            | outcome            |
|  |                | 2010-2012          | 2011-2013          | 2012-2013          |
| Income earned from paid employment (yes) in %    | Kernel, Levels | 8.4***<br>(5.42)   | 12.1***<br>(7.33)  | 11.7***<br>(6.19)  |
|  | Kernel, DiD    | 9.3***<br>(5.27)   | 13.0***<br>(5.84)  | 12.2***<br>(6.50)  |
|  | Radius, DiD    | 9.8***<br>(4.81)   | 12.0***<br>(4.62)  | 13.2***<br>(4.83)  |
| Annual income earned from paid employment in CHF | Kernel, Levels | 1,549**<br>(2.83)  | 2,313***<br>(4.16) | 3,479***<br>(4.85) |
|  | Kernel, DiD    | 1,983***<br>(3.85) | 2,747***<br>(4.95) | 3,975***<br>(5.37) |
|  | Radius, DiD    | 2,198***<br>(3.64) | 2,922***<br>(4.57) | 4,017***<br>(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 definition of the outcome variables.

## VII. Sensitivity Analysis

Since the analysis is based on the selection on observables assumption / conditional independent assumption (CIA), the question arises whether and under which conditions unobserved variables may significantly change the estimated ATT effects.<sup>7</sup> Possible candidates are the educational background or 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

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

<sup>7</sup> Since the matching approach applied conditions on pre-treatment outcome variables and additionally uses DiD, the question could be reformulated: are there *time-varying* unobserved variables which may bias the results?

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 propensity score and the calculation of the ATT effect. This procedure is repeated for 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 sizeable 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 are 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 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 like:                                     |                  |          |          |          |                |                  |                 |
| 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 effect is based on 250 repetitions.

As mentioned in Section V the bandwidth choice does not seem to be very important here. This is shown in the following 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 seem worse. (2.) The differences in the estimated ATT effects are very small. Part of the differences may be driven by treated persons being of "off support" in case of the smaller bandwidth (0.01).

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

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

| Band-<br>width | Treated:<br>off<br>support | ATT effect on monthly total DI<br>benefit in CHF (t-stat.) |                 |                 | Match quality, summary measures |                          |           |           |           |           |
|----------------|----------------------------|--|-----------------|-----------------|---------------------------------|--------------------------|-----------|-----------|-----------|-----------|
|                |                            | $t + 1$  | $t + 2$         | $t + 3$         | Pseudo-<br>R <sup>2</sup>       | Likelihood<br>Ratio Test | Mean Bias | Med. Bias | Rubin's B | Rubin's R |
| 0.01           | 5                          | -56<br>(-3.07)   | -108<br>(-4.86) | -135<br>(-5.28) | 0.003                           | 9.63                     | 1.7       | 1.5       | 13.8      | 1.03      |
| 0.03           | 1                          | -62<br>(-3.58)   | -115<br>(-5.38) | -142<br>(-6.04) | 0.003                           | 7.49                     | 1.5       | 1.1       | 12.9      | 0.97      |
| 0.06           | 0                          | -63<br>(-3.88)   | -115<br>(-5.68) | 142<br>(-6.12)  | 0.003                           | 6.90                     | 1.3       | 0.9       | 12.3      | 0.99      |
| 0.10           | 0                          | -62<br>(-4.06)   | -115<br>(-6.18) | -142<br>(-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 the persons “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 government'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 government. This is the case if the initial expenditures of around CHF 8,740 (= € 8,020 or USD 8,767)<sup>8</sup> per participant were overcompensated by 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.

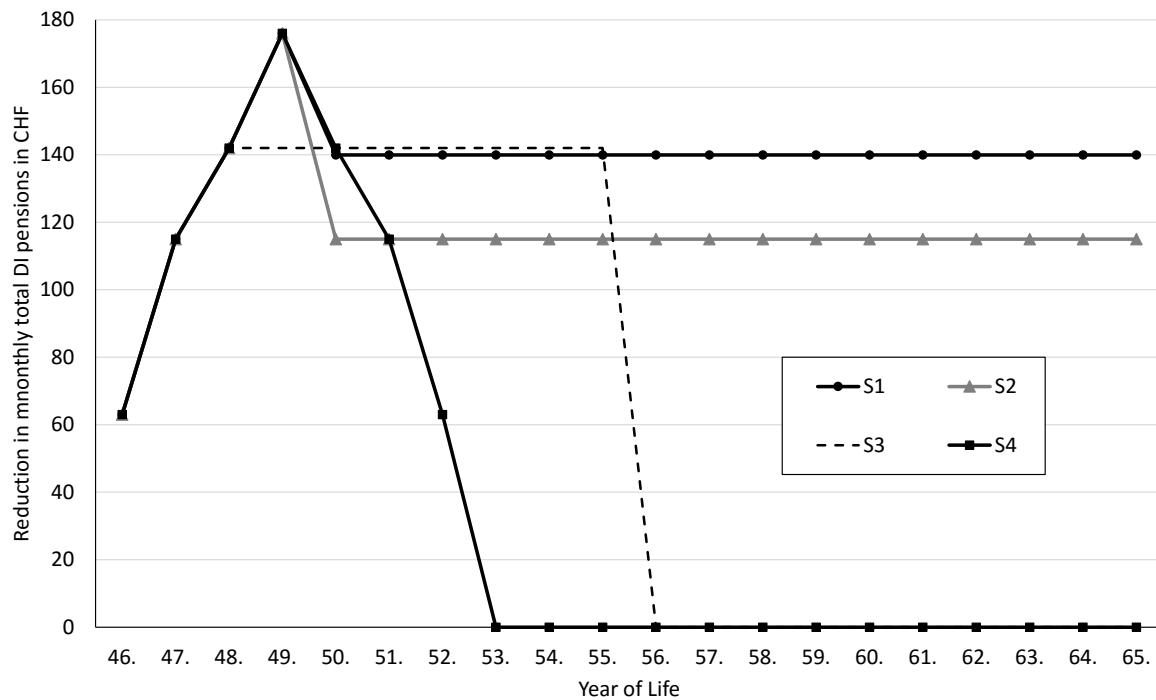
Any assessment of future reductions in payments must be based on arbitrary assumption since the empirical estimates on the ATT effects are only available for the first four years after the individual start of the program. Moreover, estimated ATT effects for  $t + 4$  are based on a reduced number of participants (578 instead of 908). However, the entire time period until the start of the old-age pension is of relevance in this context. Given the participants' average age of 45, that amounts to a 20-years period. Two assumptions already mentioned in Section IV are necessary: (1.) The treatment does not affect mortality. (2.) The treatment does not affect the retirement age. While the former assumption seems reasonable, the second assumption is problematic. However, since it is not possible to estimate the effect of the treatment on the retirement age there is no alternative to this assumption.

Various scenarios are applied in order to determine what effects the treatment would have on total DI benefits over the period until the participants 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 is 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 the 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).
- **S2**: The ATT effects are from  $t + 1$  to  $t + 4$  according to the estimates in Table 8. From  $t + 5$  until the old-age pension the ATT effects are constant at the level of the period  $t + 2$ .

<sup>8</sup> Exchange rates 28-February-2016.

- **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 7 years.



Notes:  $t + 1$  corresponds to the 46. 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<sup>9</sup> are estimated, it is possible to calculate the additional social security contributions resulting from higher income. The total contribution rate, which is assumed be constant in the future, is 12.5%.<sup>10</sup> However, due to data availability as described in Section IV, the ATT effects are only estimated up to  $t + 3$ . Hence, the

<sup>9</sup> 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 3,922 (4.91) in  $t + 3$ .

<sup>10</sup> 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 (InIVA), Switzerland

scenarios described above have to be adjusted.<sup>11</sup> While it is possible to estimate the effects on social security contribution, this is not possible with regard to the income tax (or other taxes), because not all necessary information (such as further incomes) are available.

TABLE 12: CURRENT VALUES OF REDUCTIONS IN TOTAL DI BENEFIT PAYMENTS AND ADDITION SOCIAL SECURITY CONTRIBUTIONS BY SCENARIO AND DISCOUNT RATE (PER PARTICIPANT IN CHF)

| real discount rate                        | 1.0%   | 2.0%   | 3.0%   |
|---|--------|--------|--------|
| <b>S1</b>                                 |        |        |        |
| 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 |
| <b>S2</b>                                 |        |        |        |
| 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 |
| <b>S3</b>                                 |        |        |        |
| 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 |
| <b>S4</b>                                 |        |        |        |
| 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 2.8% for the period 1981 to 2014. However, this may not be the appropriate discount rate for government programs. At present (Feb. 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 on future real interest rates. Given the low returns on Swiss bonds and the

<sup>11</sup> S1: The ATT effects are from  $t + 4$  until the 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 the 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 effects are assumed to be zero.

discussion on “Secular Stagnation” (Baldwin and Teulings 2014), 3.0% seems to be a reasonable upper bound for the average real interest rate of the time period between the measure (2009-2013) and the upcoming 20 years. 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 future reductions in DI benefits and supplementary benefits as well as the current value of future additional social security contribution are calculated.

Table 12 shows the results, which are rounded to CHF 100 in order to avoid a spurious accuracy. Comparing the governments investment of CHF 8,740 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 government. This is even more true, when summing up all three positions (DI benefit, supplementary benefits, social security contributions). Based on this the expected net benefit of the government is between CHF 8,000 up to CHF 48,400 per participant.

## **IX. Summary and Conclusions**

During 2009-2013 a pilot project was carried out in Zurich which aimed at increasing 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 which is specialized 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 possibilities to provide additional financial incentives or to make use of “threat effects”. (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 governments perspective. 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. By estimating the treatment effects for the subgroup of dropouts from the program further indications on the robustness of the results are generated. For the balancing of observable

variables kernel-based matching on the propensity score as well as radius matching with bias adjustment and additional variables in the Mahalanobis distance are applied. In terms of “balancing tests” kernel-based matching outperforms the radius matching estimator. The estimated treatment effects do not differ significantly with the method.

All the estimation results indicate that, on average, the project was effective for the participants in 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 4 years after the starting year there is no indication that this positive effect is only temporary. Based on different scenarios and different discount rates, the current value of future reductions in benefit payments and additional social security contributions is calculated. Subtracting the costs of project, the expected net benefit of the government is between CHF 8,000 (= € 7,340 or USD 8,025) up to CHF 48,400 (=€ 44,411 or USD 48,550) per participant.<sup>12</sup>

How can the 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 hamper 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 can participate in traditional VR measures. Moreover, the findings of this paper are well in line with previous studies on VR. 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.

Nevertheless, given the fact that many participants had massive health problems and that they had received DI benefits for many years when starting the program, the result that a pure placement coaching program is effective is an amazing insight. 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 measure which does not include any explicit investments in human capital, financial incentives or threat effects.

<sup>12</sup> Exchange rates 28-February-2016.

## References

- Aakvik, A., Heckman, J. J., & Vytlačil, 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 microeconomic 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.
- 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., & 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* (No. 09018). Otto-von-Guericke University Magdeburg, Faculty of Economics and Management.
- World Bank (2015), World Development Indicators, retrieved 09-Feb-2016.