Swedish Institute for Social Research (SOFI)

Stockholm University

WORKING PAPER 3/2005

JOB-SEARCH ASSISTANCE USING THE INTERNET - EVIDENCE FROM A SWEDISH RANDOMISED EXPERIMENT

by

Pathric Hägglund

Job-search Assistance Using the Internet
- Evidence from a Swedish Randomised Experiment♠

by

Pathric Hägglund*

Swedish Institute for Social Research

May 2005

Abstract

This paper reports the experience from a randomised experiment offering voluntary job-search assistance on the Internet to job seekers at Swedish public employment offices. Among those applying for participation, youth, highly educated and people living in big city areas were overrepresented. The evidence suggests that common difficulties inherent in the experimental approach, such as ethical concerns, bureaucratic behaviour and randomisation bias, have been circumvented. However, due to the voluntariness, the programme suffers from compliance problems in terms of both no-shows and drop-outs. The experimental intent-to-treat impact estimate fail to reject the hypothesis of a zero programme effect. Finally, a methodological comparison suggests that standard nonexperimental techniques succeed in reproducing the nonbiased experimental results.

Keywords: Internet job search, policy evaluation, social experiment

JEL classification: C93, J64

♠ I would like to thank Johanna Sköldung and Evy Green at the Swedish Labour Market Board (AMS) for excellent assistance setting up and executing the experiment. I am also grateful for valuable comments from Anders Björklund, Kenneth Carling, Anders Harkman and Michael Rosholm, and also seminar participants at the Swedish Institute for Social Research at Stockholm University, and the Office of Labour Market Policy Evaluation (IFAU) in Uppsala.

* Swedish Institute for Social Research, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: pathric.hagglund@sofi.su.se.

Introduction

This paper reports the experience of a demonstration programme offering voluntary job-search activities on the Internet to job seekers at Swedish public employment offices. By using random assignment to these programme services, the study contributes to the sparse literature on experimental evaluation of labour market topics in Europe in general, and in Sweden in particular. The current case in fact represents the first Swedish experiment in this field since 1975.[1]

The non-random selection of programme participants constitutes a serious threat in estimating nonbiased policy effects. Random assignment is the statistical solution to this problem, since it balances the properties affecting the subsequent outcome between members of the experimental and the control group. However, experiments can create problems of their own. A careful evaluation design is necessary to avoid biases inherent in the experimental approach.

This paper investigates i) how the experimental evaluation design succeeds in circumventing common difficulties in experimental assessment,[2] ii) the employment outcome from pursuing voluntary job-search club services on the Internet, and iii) which nonexperimental evaluation methods are likely to produce consistent results in evaluation situations similar to this one, that is, in the absence of an experimental de-

[1] To my knowledge, apart from Delander (1978) on Swedish data, the only other experiments conducted in Europe are reported in White & Lakey (1992) in the UK, Torp, Raaum, Hernaes & Goldstein (1993) in Norway; Raaum, Torp & Goldstein (1994) in Norway; van den Berg & van der Klaauw (2001) in the Netherlands; Bratberg, Grasdal & Risa (2002) in Norway; and Rosholm & Skipper (2003) in Denmark.

[2] See for instance Heckman & Smith (1995) or Björklund & Regnér (1996) for a general presentation of these difficulties.

sign. The last analysis is performed through a comparison between the presumed nonbiased experimental impact estimate and those derived from using ex-post constructed comparison groups.

Several findings can be extracted from this study. First of all, the clients´ interest in the services was lower than expected. Despite extensive advertising on all public offices and the public employment services (PES) homepage, only 636 proper applications were submitted. Second, the experimental design successfully avoids many experiment-related problems such as ethical concerns, bureaucratic behaviour and randomisation bias. However, the voluntariness of the services generated a large fraction of no-shows and early dropouts which makes it difficult to fully utilise from the experimental set-up. Third, the compliance problems combined with the small sample sizes, producing impact estimates with unsatisfactorily low precision, contributed to a low insignificant mean-difference impact estimate. Performing nonexperimental analysis, taking advantage of the large variations in the treatment dose among the experiment group members, supports a zero programme impact. Finally, testing the performance of various nonexperimental estimators, these generate impact estimates close to the mean-difference estimator. This would suggest that the available data efficiently captures the mechanisms underlying the self-selection process. However, with the small set of observations in the experiment, vast differences between the experimental and nonexperimental impact estimates had been necessary for this *not* to hold.

The following exposition is divided into two parts. The first focuses on the programme contents and the experimental design, describing in detail the virtual job-search club services and the implementation phase. Data and descriptive statistics are presented as well as experimental and nonexperimental impact estimations on the transition to employment over a six-month period. The second part, which is more technically advanced, introduces frequently applied nonexperimental estimators and their identifying assumptions. Finally, I report the performance of these estimators in reproducing the experimental results.

THE EXPERIMENT

The job-search club services

The Internet presents new opportunities for the public employment service (PES). Since 1995, several on-line placement services have been introduced in Sweden. The Vacancy bank, where employers advertise their job vacancies, had 450 000 visitors in April 2001. In 2001 the PES Internet services in Sweden were used every month by more than 550 000 individuals, which corresponds to approximately 15 per cent of the workforce.

Current developments involve a higher degree of interactivity between job seekers and employment officers, which means that further dimensions in the field of traditional employment services are being added to the Internet services. In the spring of 2002 a small committee at the Swedish labour market board (including myself as administrator of the experiment) was assigned to carry out a nation-wide demonstration programme, investigating the possibility of pursuing traditional job-search club activities on the Internet.[3] The results were to provide the basis for a policy decision as to whether or not the services should be a permanent feature of the employment services. Of particular interest was the service's ability to improve effectiveness of the matching. The services were tested on a group of voluntary job seekers who, at any time, had the opportunity to quit. The services were offered *in addition* to the regular services at the employment offices, hence, participants were subject to the same basic treatment as nonparticipants.

The job-search club services were executed by three full-time employed case workers (coaches) situated in a public office in Stockholm. No specific requirements were specified as regards the activity among the participants. They were, however, recommended to visit the programme every day. The only prerequisite as a participant, besides being registered at the employment office, was to have access to a computer with email and Internet facilities away from the local employment office. This was crucial since the participants could not access the job-search club services at their local employment offices.

In contrast to many of the more expensive labour market programmes, evaluations of traditional job-search assistance generally show positive outcomes.[4] In Sweden, both experimental and quasi-experimental evaluations conclude enhanced job chances, at least for measures targeted to subgroups of unemployed.[5] Similar to traditional job-search clubs, the concept of the virtual version was to teach job-seeking skills. The pro-

---

[3] The local labour market board of Västra Götaland first introduced the services in October 2000.

[4] See Martin & Grubb (2001) for an international review.

gramme provided guidance as to where and how to make contact with suitable employers. An important part of this was to help the participants to discover their own good qualities and to strengthen their self-confidence. Participants learned how to write job applications and CVs and how to behave during job interviews. The theoretical elements were combined with practical exercises whereby the participants received feedback from the coaches. The programme allowed interactivity among programme members and the benefits of group dynamics. Organised group discussions and on-line chats were permanent features of the services.

A comparison between the traditional and virtual version of the job-search club services also reveals some important dissimilarities. First of all, participants in the former receive services according to a predetermined schedule supervised by caseworkers. The activities in the latter involve working in an Internet environment, where the participants choose for themselves when, where from, and for how long they wish to be active. Secondly, instead of participants working in the presence of caseworkers and other participants, they are expected to work individually and away from the employment office itself.

Experimental design

The virtual job-search club activities were carried out in the summer of 2002. To evaluate the impact from the services, voluntary job seekers at the employment offices

---

[5] See Calmfors, Forslund & Hemström (2001) for a review of Swedish experiences.

were randomly selected into two groups. One of the groups was, in addition to their regular services, offered the job-search club services.[6] The other group was directed to the regular services. The access to this extra assistance was expected to have a positive impact on the participants´ job chances.[7] Anyone currently registered as a job seeker was welcome to apply for participation. This included openly unemployed, programme participants as well as employed persons looking for a new job. Furthermore, since the services were offered on the Internet, no geographical restrictions were introduced.

The programme ended on 6 September 2002. For those offered the services, an early entry date allowed approximately three months' services. The control group was denied programme access for a further three months. The success indicator here is "exit to employment" during the six-month follow-up period (between 15 May /5 June and 1 December 2002). In constructing this indicator, special attention is paid to those already employed at the start of the experiment. Only employment resulting in a move "up-wards" in the ranking system counts as a successful outcome.[8]

An alternative choice of outcome measure, also presented here, is employment status 1 December 2002. Compared to the cumulative exit indicator, this measure adds a

---

[6] Note that the services then need to provide positive programme effects in order to be economically motivated.

[7] From a (traditional) job-search theory point of view, two opposite effects are expected on job transitions (and hence on unemployment duration). First of all, the job offer arrival rate is expected to increase. Second, from more job offers follow a higher degree of selectivity in choosing which offer to accept, i.e., a higher reservation wage. The net effect is thus ambiguous. However, van den Berg (1994) shows that the positive effect dominates the negative effect under weak restrictions on the wage offer distribution.

[8] The following employment-type ranking, based on unemployment register information, is applied to those already employed: 1) Regular employment, 2) Job-changer, 3) Temporary employed, 4) Part-time employed or employed by the hour. If a person exits to employment involving a higher employment-type ranking, the person counts as employed. If a person remains in the same employment-type category, or

quality dimension to the employments taking into account the potential flow back to unemployment. However, since the unemployment register data (Händel) does not include information on employment status, we have to presuppose that the current employment status among those who exited from unemployment, and did not return, is equal to the cause of separation. Considering that the sample includes students looking for a temporary occupation, this indicator overestimates the true employment rate. Furthermore, specifying an outcome variable as the status at a certain date makes the result sensitive to the particular date chosen. A third possible measure comparing the performance of experiment and control group members is unemployment duration. This approach, however, involves relatively sophisticated analytical methods which would interfere with the otherwise simple and intuitively understandable feature of the experimental approach.

Although application for participation was voluntary, most of those offered the programme services did not immediately take action. To encourage participation, applicants were contacted by email and/or telephone, and were reminded of the service offer. They were also told that their password would expire on a certain date. Although this increased the participation rate, a significant fraction of the experimental group never took part in the activities at all. This is discussed further in the next section.

A total of 843 valid applications were received in two enrolment periods (Table 1). The first, in which information about the services was available at all local employment offices in Sweden, took place between 29 April and 10 May. The second enrol-

---

exits to an employment with a lower ranking, the opposite holds. Note that people in the highest ranked ment period, in which applications could be submitted either at the employment offices or on the PES homepage, took place between 21 May and 31 May.

Table 1. Sampling scheme

	First enrolment (April 29-May 10 2002)[a]	**Second enrolment** (May 21-May 31 2002)[b]	**Total**
No. of applicants	346	497	843
of whom:			
Registered at the employment office	**265**	**371**	**636**
of whom:			
Experimental group	**140**	**203**	**343**
of whom:			
Participants	68	113	181
No-shows	72	90	162
Control group	**125**	**168**	**293**

Note: Corresponding start dates, [a]: 15 May, and [b]: 5 June.

Fortythree applications were eliminated either because an invalid email address was given or because the applicant had found a job before the start date. Another 164, not currently registered at an employment office, were also excluded. Of the remaining 636 job searchers, 343 were randomised into the experimental group that was offered services. Of these 181 (here, “participants”), or 53 per cent, visited the job-search programme home page at least once, while 162, or 47 per cent, never visited at all (here, “no-shows”). The control group, 293 persons, did not receive any service offer during the follow-up period, but were directed instead to the regular services of their local em-

---

category are not included constructing the sample.

ployment offices. The applicants were informed by email whether or not they were to be admitted to the experimental programme at the start dates, i.e., 15 May or 5 June 2002.

Common difficulties in experimental evaluation

New ways of organising activities are particularly suitable for experimental evaluation. Björklund & Regnér (1996) conclude: “Indeed, we are convinced that this (*read: alternative ways of organising job-search activities*) is the field where the benefits of classical experiments are the greatest and where the traditional problems can be handled most easily”.

The evaluation of demonstration programmes relating to new services, rather than established ones, offers a fairly straight forward example of social experimentation. Since in the present case, for instance, no-one was being denied services they would otherwise have been entitled to, there was no need for concern on ethical grounds that services were being denied to part of the eligible population. To oppose random assignment in such a situation implies that the relevant services should be implemented immediately without being tested first. Because nobody knows for sure that the experimental group members actually gain anything from their participation, there is no ethical reason for preferring this alternative. As is typical of small-scale demonstration programmes, there were more eligible applicants than available programme slots. Thus randomisation is not an unfair selection instrument.

Co-operation from the administrators at different levels is crucial in conducting a successful experiment. The administrators should behave as if the services were in normal operation. This requirement was most likely fulfilled although we cannot completely rule out the possibility that the coaches were too enthusiastic about the programme, and therefore more effective than under normal conditions. However, avoiding ethical concerns most certainly had a positive impact on the willingness among programme administrators to cooperate and to follow the outlined evaluation strategy. The demonstration programme thus reduced the risk of bureaucratic resistance.[9]

Also, evaluating new types of work organisation and using new technology, almost by definition eliminates the risk of *substitution bias* that occurs when control-group members receive services similar to those being offered to the experimental group. In the case of these job-search club services, there were no obvious substitutes. Also, since the local administrators had no chance of controlling the assignment process and possibly distort the experiment group, and since the applicants were not told that the programme was being evaluated, bias due to a nonrepresentative pool of participants (*randomisation bias)*, and/or to participants altering their behaviour during the programme (*Hawthorne effect*), could be ruled out. Finally, by not imposing geographical constraints, it was possible for even a small-scale programme to be carried out nationwide. Hence, the risk of *displacement effects* due to experiment group members acquiring employment at the expense of control-group members, was significantly reduced.

While the design manages to avoid several typical problems inherent in social experiments, some important topics still remain. First of all, as shown in Table 1, the experiment involves relatively few observations (636). This suggests that the pro-

[9] However, we cannot completely rule out the possibility that the coaches were too enthusiastic about the

gramme impact estimators will produce estimates with low precision, which suggests only very large effects will have a chance of becoming statistically significant. The small-scale dilemma is present in several of the European experiments. For example, the Berg & van der Klaauw experiment (2001) included 394 UI receivers, the Bratberg, Grasdal & Risa study in 2002 was based on a sample of 560 workers on sick leave, and the Rosholm & Skipper paper from 2003 contained 812 unemployed applying to participate in labour market training programmes. Also, the 1975 Swedish Delander study consisted of 410 currently unemployed.

Second, Table 2 presents three different measures of the experiment group members' level of activity in the job-search club. The measures reveal both large proportions of no-shows, i.e. people who never entered the programme, and dropouts, i.e., people who dropped out of the programme prior to receiving all of the treatment. The first column shows that almost 50 per cent of the experiment group never visited the job-search programme home page. Of those who did, 40 per cent did it on one occasion only. The second column tells us that only about 30 per cent actively used the services in more than one hour. According to the third column, between 70 and 80 per cent of the experiment group members failed to complete any of the exercises. The presence of no-shows and dropouts dilutes the experimental estimator because the difference in treatment between the experiment and control group is reduced. The compliance problem is well documented. Heckman et al. (1999) shows that in experiments conducted in the U.S between 1975 and 1992, the portion of experiment group members receiving

---

programme, and therefore more effective than under normal conditions.

treatment was often less than 0.7, in some cases even below 0.5. In Europe, the Rosholm & Skipper experiment suffers from both no-shows (48%) and cross-overs (22%), i.e. control group members receiving treatment.

There could be several reasons for the large amount of no-shows in this experiment.[10] The services being nonmandatory is most likely one important explanation. Just as submitting an application was voluntary, so was participation. No penalty was imposed on those who ignored the possibility of joining the programme. The lack of computer availability could be another explanation. Although specified as a prerequisite, not all applicants would necessarily have had access to an outside computer. Finally, some of the absence could be due to deficiencies in data. Using information on current unemployment spells from the unemployment register, data to a large extent is based on self-reported information. For instance, when registered job seekers find employment they could omit to report to the employment office. As a result, the register would overestimate the true number of unemployed at any given moment. Hence, although the experiment and control groups at the programme start date consisted of currently registered job seekers, they possibly include persons no longer unemployed.

Similar to the case of no-shows, there are potentially multiple reasons for the presence of dropouts. Clearly, the voluntariness allowed participants not fully satisfied with the services to quit. However, the services encouraged practising the skills during

[10] It is important to distinguish between no-shows and “attriters”. No-shows do not receive the services but remain in the follow-up sample, while attriters are usually eliminated. In our sample, 10 per cent in the experimental and 10 per cent in the control group were deregistered and coded “cause unknown” (attriters). This indicates that the employment officer lost contact with the unemployed. Since the attrition is not systematically related to either of the groups, the attriters are *not* excluded from the sample.

treatment, which means that getting a job is one likely reason for not pursuing the programme. Furthermore, as displayed in Table 3 below, the services particularly appealed to the group of young job seekers. More than 25 per cent of the applicants were below the age of 25. This is generally a mobile group of unemployed with, on average, short spells of unemployment. The group is also highly prioritised by the authorities which means that their unemployment spells are more frequently interrupted by active measures. Finally, the chosen period of performing the services, including the summer months June, July and August, is probably an additional explanation for the small-dose problem. In these months, the search activity is generally lower on average.

In sum, several common pitfalls associated with experimental and even nonexperimental evaluation have been avoided here. The voluntariness, however, most likely contributed to an imperfect experiment where the majority of those offered the services either denied the offer, or dropped out early. This would probably have been avoided had the services instead been a compulsory full-time activity. On the other hand, then other problems, for instance ethical objections, bureaucratic behaviour and randomisaton bias, would potentially have been issues of more concern. An agreement of some sort would perhaps have increased the compliance intensity. But then a sanction system would have been necessary to maintain these agreements. To the extent that some of the no-shows were caused by the applicants failing to fulfil the requirements as participants, these could have been minimised had the randomisation and programme start been preceded by an outreach procedure. Then the computer availability criterion, as well as the job seeker status prerequisite, could have been confirmed.

Table 2. Distribution of three measures of activity in the job-search club among experiment group members

Percentile	**No. of accessions**	**No. of operative minutes**	**Per cent of exercises completed**
0	0	0	0
10	0	0	0
20	0	0	0
30	0	0	0
40	0	0	0
50	1	27	0
60	1	36	0
70	1	65	0
80	4	183	6
90	11	520	35
100	327	4539	100
Mean	*6*	*200*	*8*

Note: Number of observations: 343.

The evaluation problem

The fundamental evaluation problem arises because a person cannot be observed in two labour market states at the same time. Consequently, the evaluation problem is typically formulated at the population level and focuses on mean impacts of participation. Using similar notation as Heckman et al. (1999), let $D=1$ indicate the offer to participate in the programme, $D=0$ otherwise, and Y_1 and Y_0 the respective outcomes. The average treatment effect on the treated is then:

$$E(\Delta \mid D=1) = E(Y_1 - Y_0 \mid D=1) = E(Y_1 \mid D=1) - E(Y_0 \mid D=1). \quad (1)$$

In reality, we observe Y_1 for those treated and Y_0 for the nontreated. Comparing means between the observables we get:

$$E(Y_1 \mid D=1) - E(Y_0 \mid D=0) = E(Y_1 \mid D=1) - E(Y_0 \mid D=1) + \{E(Y_0 \mid D=1) - E(Y_0 \mid D=0)\}, \qquad (2)$$

which equals the average treatment effect on the treated plus a bias term. The last part of equation 2 is attributable to the fact that the outcomes of those not offered treatment are not necessarily representative of the nonobservable outcomes of those offered treatment had they not been offered it. Given that treatment is randomly assigned, the selection-bias problem is solved because D is independent of the potential outcomes. As a consequence, the bias term within braces in equation 2 equals zero and

$$E(Y_1 \mid D=1) - E(Y_0 \mid D=0) = E(\Delta \mid D=1) \,. \qquad (3)$$

Random assignment thus ensures that all those offered treatment and all those not offered treatment are comparable as groups, and that differences in the subsequent outcomes are attributable to programme participation. If, however, for some reason members of the experiment group fail to receive treatment (no-shows), equation 3 no longer fulfils the requirements of a treatment-on-the-treated estimator. Since selection into participation is expected to be non random, and since it is impossible to identify the corresponding participants in the control group, we can no longer assume that the entire control group meets the requirements for a suitable counterfactual. Rather than the treatment effect on the treated, the experimental mean-difference estimator estimates the effect of the *availability* of the services, or the *intent to treat*. This is, however, also a policy-relevant parameter because it provides information about how the receipt of an offer to participate affects the subsequent outcome.[11]

Data and descriptive statistics

The experiment and control group members have been followed in Händel, an event database administered by the Swedish Labour Market Board. Händel records all unemployment and labour market programme (LMP) periods, as well as the causes of separation, since August 1991. The register contains information about *personal characteristics* (gender, age, educational level, citizenship, working disability, community etc), and *profession* (desired profession, experience and education in desired profession). The longitudinal character of Händel also makes it possible to define variables reflecting an individual's *unemployment history* (for instance, duration of the ongoing unemployment spell, total duration of all unemployment spells, number of times openly unemployed and number of LMPs embarked upon).

Studying the descriptive data of the various groups of job seekers in Table 3, three different comparisons are especially interesting. First of all, with this experiment

[11] In order to recover the treatment effect on the treated, we must adjust equation 3 for the presence of no-shows. If T is introduced as an indicator of programme services actually being received, where $T=1$ represents participation and $T=0$ otherwise, then $\frac{E(Y_1 \mid X, D=1) - E(Y_0 \mid X, D=0)}{P(T=1 \mid X, D=1)}$ resolves the treatment effect on those treated. The equation simply scales up the mean-difference estimate by the fraction of participants in the experimental group. To estimate the treatment impact correctly, one assumption is that the mean outcome of no-shows in the experimental group is the same as their analogs in the control group, that is; $E(Y \mid X, D=1, T=0) = E(Y \mid X, D=0, T=0)$. Note that in presence of dropouts, the treatment-on-the-treated estimator more accurately represents various levels of partial treatment, rather than the effect of full treatment.

being the first test of the clients´ interest in the services, it is interesting to examine what characterise those who applied for participation. A comparison between all job seekers (column 1) and the full experimental sample (column 6), shows women to be significantly underrepresented among the latter (t-tests in column 7). This could be due to the relatively large proportion of women registered at the employment offices as part-time employed. Since employed persons, as well as programme participants, are less attracted by the services, this is a natural consequence. Rather expectedly, the applicants are somewhat younger and more highly educated. The differences are especially noticeable for those in the 18-24 and 25-34 age range, and among those with experience from university studies. This last corresponds to the results presented in Kuhn & Skuterud (2002). Accordingly, people seeking jobs demanding special theoretical competence or shorter university education are overrepresented among the applicants. Compared to the average job seeker, the applicant group also contains a larger proportion of people living in big cities, and people experienced and educated in their desired professions. Finally, the applicants have initiated a larger number of unemployment spells and have more often participated in programmes. They are also currently experiencing unemployment spells that are only half as long as those of the non-applicants.

A second comparison, answering to the question of how successful the randomisation was, is between the characteristics of the members of the experimental and control group. Except for random differences, the groups should be similar regarding both observables and nonobservables. The random differences diminish with the number of observations. In small samples, however, the discrepancies can be quite substantial. Comparing the groups in columns 4 and 5 reveals that the mean deviations are almost exclusively small. However, 31 per cent in the experimental group had received at least two years of higher education (university), compared to only 24 per cent in the control group. The difference is statistically significant at the 5% level (t-tests in column 8). Since educational level is usually positively correlated with employment probability, the simple mean-difference estimator could overestimate the true programme effect. The experiment group also comprises a significantly larger proportion of job seekers looking for craftsman´s work. Finally, members of the experimental group had started on significantly fewer LMPs than the control-group members at the experiment start. Note that with 54 variables, the groups would be expected to significantly differ in 2-3 of those (0.05*54=2.7).

Comparing participants and no-shows among those offered services (columns 2 and 3), reveals a non-random selection into participation. The no-show rate is higher in the youngest age category and among those with a low educational level, as opposed to the age category 25-34 and job seekers with more than two years of university studies. Also, people who are currently employed or taking part in a programme more often reject the offer of participating. An interesting result is that the no-show rate is higher among those who submitted their applications at the employment offices (54%), rather than through the PES homepage (28%). Since the choice of application channel may signal whether or not an individual has access to an outside computer, this is a useful finding.

Table 3. Summary statistics for the experimental group (participants and no-shows), the control group and all registered job seekers. Bold type indicates statistical significance at <5%-level.

	All job seekers (1)	**Experimental group**			**Control group (5)**	**Full experi-ment sample (6)**	**t-test (1)-(6) (7)**	**t-test (4)-(5) (8)**
		Participants (2)	*No-shows (3)*	*Mean (4)*				
Gender								
Female	0.52	0.43	0.40	0.41	0.45	0.43	***4.40***	*-0.84*
Male	0.48	0.58	0.60	0.59	0.55	0.57	***-4.40***	*0.84*
Age								
18-24	0.18	0.22	0.34	0.27	0.26	0.27	***-4.98***	*0.42*
25-34	0.25	0.35	0.29	0.32	0.33	0.32	***-3.64***	*-0.19*
35-44	0.24	0.22	0.24	0.23	0.22	0.22	*1.19*	*0.27*
45-54	0.19	0.16	0.11	0.14	0.14	0.14	***3.11***	*-0.23*
55-	0.14	0.06	0.02	0.04	0.05	0.05	***10.21***	*-0.62*
Mean	*38.2*	*34.5*	*31.6*	*33.1*	*33.5*	*33.3*	***10.82***	*-0.38*
Educational level								
<Compulsory school	0.11	0.03	0.07	0.05	0.07	0.06	***4.82***	*-0.99*
Compulsory school	0.19	0.12	0.17	0.14	0.17	0.16	***2.24***	*-1.17*
Upper secondary	0.5	0.36	0.45	0.4	0.43	0.42	***4.09***	*-0.79*
University <2 years	0.06	0.1	0.09	0.09	0.08	0.08	***-2.32***	*0.83*
University >=2 years	0.14	0.39	0.23	0.31	0.24	0.28	***-7.50***	***2.07***
Graduate level	0.00	0.01	0.00	0.00	0.01	0.01	*-0.69*	*-1.11*
Home county								
Big city[a]	0.42	0.59	0.60	0.60	0.56	0.58	***-7.82***	*0.80*
Local labour markets[b]	0.26	0.16	0.11	0.14	0.13	0.13	***8.51***	*0.27*
Other	0.32	0.25	0.29	0.27	0.31	0.29	*1.95*	*-1.08*

Experience in desired profession								
No	0.35	0.28	0.28	0.28	0.32	0.30	***2.87***	*-1.12*
Yes	0.65	0.72	0.72	0.72	0.68	0.70	***-2.87***	*1.12*
Education in desired profession								
No	0.53	0.37	0.45	0.41	0.39	0.40	***6.46***	*0.50*
Yes	0.47	0.64	0.55	0.6	0.61	0.60	***-6.46***	*-0.50*
Employment type in desired profession								
Full-time	0.51	0.55	0.51	0.53	0.5	0.52	*-0.53*	*0.73*
Part-time	0.08	0.02	0.02	0.02	0.03	0.02	***7.53***	*-1.07*
Full-time/part-time	0.42	0.44	0.47	0.45	0.47	0.46	***-2.00***	*-0.40*
Desired profession								
No classified profession	0.16	0.05	0.07	0.06	0.08	0.07	***7.75***	*-1.00*
Management work	0.01	0.03	0.01	0.02	0.01	0.01	*-0.59*	*1.50*
Special theoretical competence	0.07	0.27	0.09	0.19	0.16	0.18	***-7.10***	*0.75*
Short university education	0.08	0.18	0.08	0.13	0.15	0.14	***-3.86***	*-0.56*
Administrative work	0.12	0.09	0.14	0.11	0.17	0.14	*-1.27*	*-1.93*
Service, health care and commercial work	0.23	0.13	0.25	0.19	0.22	0.20	*1.32*	*-0.90*
Farming, forestry and fishing	0.02	0.01	0.01	0.01	0.01	0.01	***4.59***	*-0.16*
Craftsman's work	0.11	0.08	0.14	0.11	0.06	0.09	*1.79*	***2.42***
Machine work, transport and communication	0.11	0.09	0.09	0.09	0.08	0.08	***2.07***	*0.41*
No vocational training required	0.10	0.07	0.12	0.09	0.07	0.08	*1.32*	*0.99*
UI-compensation								
Non	0.17	0.17	0.29	0.22	0.26	0.52	***-3.91***	*-1.12*
Base premium	0.08	0.16	0.17	0.16	0.19	0.02	***-6.17***	*-0.91*
Income-related	0.75	0.68	0.54	0.61	0.55	0.46	***8.10***	*1.68*
Working disability								
No	0.80	0.92	0.88	0.90	0.88	0.89	***-6.97***	*0.94*
Yes	0.20	0.08	0.12	0.10	0.12	0.11	***6.97***	*-0.94*
Citizenship								
Swedish	0.90	0.83	0.82	0.83	0.86	0.84	***4.12***	*-1.09*
Other Nordic countries	0.02	0.03	0.03	0.03	0.02	0.03	*-1.33*	*0.71*
Other	0.08	0.14	0.15	0.15	0.12	0.14	***-3.86***	*0.85*

Expanded search area[c]								
No	0.88	0.75	0.83	0.78	0.79	0.79	***5.26***	*-0.23*
Yes	0.12	0.25	0.17	0.22	0.21	0.21	***-5.26***	*0.23*
Unemployment history								
No. of LMPs	3.96	4.30	5.02	4.64	5.41	4.99	***-5.47***	***-2.10***
No. of periods openly unemployed	6.61	8.12	9.02	8.54	9.49	8.98	***-9.35***	*-1.92*
Unemployment duration								
Ongoing unempl. period, years	2.06	1.06	0.91	0.99	1.10	1.04	***12.95***	*-0.80*
All unempl. periods, years	4.93	4.13	4.09	4.11	4.38	4.23	***4.04***	*-1.37*
Status at the experiment start								
Openly unemployed	0.28	0.63	0.45	0.55	0.5	0.52	***-11.88***	*1.18*
In job	0.33	0.09	0.13	0.11	0.12	0.12	***15.29***	*-0.47*
In LMP	0.39	0.28	0.42	0.34	0.38	0.36	*1.27*	*-0.91*
Start date								
15/5-02	-	0.38	0.44	0.41	0.43	0.42	*0.48*	*-0.47*
5/6-02	-	0.62	0.56	0.59	0.57	0.58	*-0.48*	*0.47*
Application channel								
ams.se	-	0.36	0.15	0.26	0.31	0.28	-	*-1.24*
Employment Office	-	0.64	0.85	0.74	0.69	0.72	-	*1.24*
Cumulative exit to employment up to 1/12-02	**0.29**	**0.37**	**0.29**	**0.33**	**0.31**	**0.32**	-	**-**
Number of observations	6899	181	162	343	293	636		

Notes: 1. “All job seekers” in col. 1 refers to a cross-section of all those registered at the employment offices on 15 May 2002. The number of observations (6,899) corresponds to approximately 1 per cent of the populationcol. [a]: Refers to the counties of Stockholm, Västra Götaland and Skåne. [b]: Refers to the counties of Värmland, Dalarna, Gävleborg, Jämtland, Västernorrland, Västerbotten and Norrbotten. [c]: During the first 100 days of unemployment, a job seeker is allowed to restrict the search area geographically.

Results

Table 4 presents the simple mean-difference estimator comparing the cumulative transitions to employment for experiment and control group members in the six months follow-up period (July-December). Comparing the full sample of experiment and control group members, the assessment is performed on the basis of the intent-to-treat principle. Both adjusted and unadjusted impact estimates are introduced. The adjusted estimator is derived by first identifying a common range in which both experiment and control group members have an actual chance of receiving an offer to participate.[12] After eliminating those lacking support, i.e. those with no counterpart in the opposite group, a Probit model adjusts the programme effect for random differences in observable characteristics.

According to the unadjusted impact estimates in the first column, the experiment group members get jobs at a somewhat higher rate, especially during the first two months. The July estimate is weakly positive significant. An explanation could be that those who were motivated to invest in the programme, were active from the start. As they gradually dropped off, for instance to employment, the average activity level in the programme diminished. This is further accentuated by the stated last possible programme start date which made late entries impossible. Hence, a programme effect, if any, would be expected to appear early in the follow-up period.

[12] The offer probability was estimated using a Probit model including the explanatory variables in Table 3, except for "Application channel". Five members of the experiment group lacked upper tail support in the control group. In the same way, four control-group members fell below the experiment group range of support.

The (unadjusted) six-month result in Table 4 is slightly positive (1.9 percentage points) but insignificant. The standard errors indicate a large 95 per cent confidence interval spread in the six-month impact estimate, from –5.4 to 9.2 percentage points. Compared to the unadjusted impact estimates, the adjusted estimates are throughout somewhat lower. The six-month effect is negative, -1.3 percentage points.[13,14] The results from the full model estimation are found in the first column of Table 6.

[13] Controlling for observables normally reduces the confidence interval surrounding the impact estimate, thus allowing smaller deviations in outcomes between experimental and control group members to become statistically significant. However, due to the loss of statistical degrees of freedom, the standard error could also become somewhat larger.

[14] Instead using *employment status at 1 December 2002* as the dependent variable, the simple mean-difference estimator generates a negative (-2.3 percentage points) effect. Probit adjustment further emphasises the negative effect, generating a point estimate significant at the 10% level (-6.1 percentage points). Combined with the results of the main analysis, this implies that the duration of the employment was on average somewhat shorter among the experiment-group members.

Table 4. Treatment effects (unadjusted and adjusted) calculated as differences in means in the cumulative exit to jobs

Month	**Differences in means (unadjusted)**	**95% conf. interval**	**Differences in means (adjusted)**	**95% conf. interval**
July	0.040*	-0.008 – 0.088	0.033	-0.008 – 0.073
	(0.024)		(0.021)	
August	0.031	-0.023 – 0.086	0.026	-0.026 – 0.078
	(0.028)		(0.026)	
September	0.007	-0.055 – 0.069	-0.006	-0.067 – 0.054
	(0.032)		(0.031)	
October	0.032	-0.037 – 0.100	0.010	-0.061 – 0.080
	(0.035)		(0.036)	
November	0.023	-0.048 – 0.094	-0.007	-0.082 – 0.068
	(0.036)		(0.038)	
December	0.019	-0.054 – 0.092	-0.013	-0.090 – 0.064
	(0.037)		(0.039)	

Notes: Standard errors, calculated as the root of $(\mathrm{var}_1/n_1 + \mathrm{var}_2/n_2)$, within parentheses. Number of observations (unadjusted): 636 (343 experiment group members and 293 control group members). Number of observations (adjusted): 627 (338/289). Adjustment based on a Probit regression estimating employment probability. The full model estimation includes the regressors in Table 3, except for information on "education in desired profession" and "application channel". 9 observations were eliminated due to a common-support restriction based on the probability of being offered the services. * refers to significance at the 10 per cent level.

Table 2 presented different measures of activity in the job-search club and revealed large variations in the treatment dose within the group of experiment group members. This variation can be explored analysing programme effects for the various treatment doses. Note that with a presupposed non-random selection into frequent and non-frequent usage of the services, such an analysis does not utilise from the random assignment. Instead a Probit-model must include conditioning variables to adjust for differences in characteristics of importance to the outcome. Still, the results need to be reserved for bias due to unobserved heterogeneity. For instance, the frequent users are potentially more "work-motivated" on average. However, with the services being offered during an extensive period of time (3 months), recurrent usage of the services could also signal problems finding employment.

The model is derived as follows:

$$Y_i^* = \beta' X_i + \gamma D_i + \varepsilon_i, \tag{4}$$

where Y_i^* is an unobservable variable related to the binear observable variable Y_i in the following way:

$$Y_i = 1 \text{ if } Y_i^* > 0, \; 0 \text{ otherwise.} \tag{5}$$

Y_i indicates whether or not an individual was employed. β is a vector of parameters corresponding to the exogenous covariates in X. As before, D_i specifies whether or not a person was offered programme services, $D_i = 1$ and $D_i = 0$. γ represents the average intent-to-treat programme impact. The error term ε_i is normally distributed.

Table 5 reports programme effects for different levels of activity in the programme. Zero accessions and zero hours of time invested represent the benchmark outcomes for each marginal dose-response impact estimate.[15] Overall, the presented estimates are insignificant and close to zero. In general, small doses of activity produce modest positive effects, with 2-5 accessions generating the largest point estimate, 5.2 percentage points. On the opposite, relatively large doses are associated with negative responses, for instance, more than 21 accessions produces a negative impact estimate of

[15] The full model estimation results are available on request.

9.3 percentage points. This possibly confirms the problem of adverse selection in the estimations.

Table 5. Dose-response Probit estimates of no. of accessions and operative hours in the job-search club

No. of accessions	Coefficient (std.err)	**No. of operative hours**	Coefficient (std.err)
1	0.028 (0.060)	**<1**	0.033 (0.063)
2-5	0.052 (0.081)	**1-1.99**	-0.006 (0.092)
6-20	-0.011 (0.081)	**2-9.99**	0.017 (0.074)
21-	-0.093 (0.097)	**10-**	0.017 (0.074)
Average	-0.002 (0.002)	**Average**	-0.004 (0.003)

Note: The sample contains 181 participants (people who entered the programme home page at least once), and 455 nonparticipants. Estimationa include the regressors in Table 3, except for information on "education in desired profession" and "application channel"

The severe compliance problems in this experiment makes it difficult to draw conclusions about the performance of the services evaluated. With the large amount of job seekers either not showing up for the services, or dropping out at an early stage, an outcome comparison between the full set of experiment and control group members becomes less meaningful. A perhaps more appropriate strategy would then be to abandon the experimental design and the random assignment of services, on behalf of the nonexperimental estimators and the non-random selection into participation. Exploring the large variation of invested time in the programme, using a simple Probit model, does not, however, provide any evidence of increased (or decreased) job chances as a result of more frequent usage of the services. However, one can not exclude the possibility that the services have a favourable impact on certain subgroups of unemployed. The results show some signs of small positive short-term, and small-dose, effects. Unfortunately, with the small samples separate subgroup analyses are not meaningful. According to Table 3, openly unemployed job seekers between the age of 18 and 34, with experience from university studies, and situated in one of the big cities, are among those most willing to test the services. Targeting to this group could thus be an idea in future investigations of the services.

Finally, if the services not being enforced made it difficult to perform a fair assessment, it clearly made the demonstration programme a relevant test of the clients´ interest in this type of services, and their expectations of the effects of the programme contents. The small-sized sample and the small-dose problem is therefore useful and policy relevant information in itself.

Evaluating Nonexperimental Evaluation Methods

Next, we use the experimental results as a benchmark in assessing the case for various nonexperimental evaluation techniques. The purpose is to examine the extent to which the available data and standard econometric methods succeed in replicating the assumed nonbiased experimental results. The method of utilising experimental data to evaluate the performance of various nonexperimental evaluation techniques has been applied in several studies, almost all of them using U.S. data. For instance, both

LaLonde (1986) and Fraker & Maynard (1987) showed that traditional econometric methods often fail to repeat experimental results. Recent nonexperimental evaluation literature places greater emphasis on matching procedures. Applied to high quality data, Heckman, Ichimura & Todd (1997) conclude that, compared to standard regression methods, these estimators generate results more consistent with those produced from experimental evaluation.[16] In Europe, a recent Norwegian study (Bratberg, Grasdal & Risa, 2002), good correspondence in the outcomes is shown when comparing experimental and nonexperimental estimators.

Nonexperimental estimators

The first model to be tested is the above presented Probit specification. This model is based on the identifying assumption that X fully captures the mechanisms affecting both the probability of receiving an offer to participate, i.e. the decision to apply, and the outcome in the absence of the programme. This is a strong assumption. If some selection is on unobservables, sample selection bias arises because

$$E(\varepsilon_i \mid X_i, D_i = 1) - E(\varepsilon_i \mid X_i, D_i = 0) \neq 0 . \qquad (6)$$

The second model to be evaluated is a variant of the familiar Heckman selection model adjusted to allow for a binear outcome variable. By extending the Probit model into a

[16] Other similar studies include: Friedlander & Robins (1995), Dehejia & Wahba (1999, 2002), Heckman, Ichimura, Smith & Todd (1998), Smith & Todd (2000).

two-equation model allowing correlated disturbances, the underlying regression relationship describing the endogenous decision to apply for participation is

$$D_i^* = \delta' Z_i + u_i \qquad \text{where } D_i = 1 \text{ if } D_i^* > 0, \text{ 0 otherwise.} \qquad (7)$$

In the equation, D_i denotes an individual's decision to apply for participation, Z_i is a vector of explanatory variables affecting the choice of submitting an application, and δ_i is a vector of coefficients. Assuming joint normally distributed error terms ε_i and u_i, equations 4 and 7 may be estimated by maximum likelihood. The selection equation preferably includes an exclusion restriction, i.e. a variable that is not part of X. In order to qualify, such a variable should be correlated with the probability of applying for participation and uncorrelated with the outcome, i.e. employment probability.[17] Here, the bivariate Probit model uses the binary variable "education in desired profession" as an exclusion restriction.

Matching procedures have become more popular in recent years, alongside the traditional parametric methods. Matching methods pair participants (in our case people receiving an offer to participate) with nonparticipants who are similar as regards observed attributes and estimate programme effects by comparing mean outcomes. However, rather than matching on a set of covariates X, Rosenbaum & Rubin (1983) showed that matching on the probability of participation, or the propensity score, also generates consistent estimates. Since finding a comparison group member becomes increasingly difficult for every covariate added in X, this is a major advantage, because the propensity score $P(X)$ is a (one-dimensional) scalar. For matching methods to properly estimate the programme impact, it is necessary that the outcome in the absence of the programme service, conditional on a set of explanatory variables, is independent of treatment T. For this *conditional independence assumption* to hold, all variables affecting both participation and nonparticipation outcomes must be observed and accounted for. Needless to say, the credibility of the matching estimator hinges on the richness of the available data. The benefit of adopting the matching approach rather than just running a simple regression, which is also conditioned on a set of observables, is first of all that matching precludes any assumptions of functional form. According to Dehejia & Wahba (2002) and Smith & Todd (2000), its non-parametric character could considerably reduce bias in the impact estimate. Another advantage is that matching methods only match programme and comparison-group members in the range of $P(X)$ that is common to both groups. Matching thus avoids comparing the incomparable.

The propensity score allows for the specification of several different estimators. The most common is the nearest-neighbour estimator, whereby participants and nonparticipants who are closest in terms of $P(X)$ are matched. This is the third estimator to be evaluated. An alternative to one-to-one matching models is models including several nearest neighbours, whereby the participant outcome is contrasted with a weighted average of outcomes. Therefore, the fourth model to be tested is a kernel-based matching model where the weight allotted to each non-treated unit is in proportion to its closeness

[17] Exclusion restrictions improve identification, although they are not formally required in parametric sample selection models.

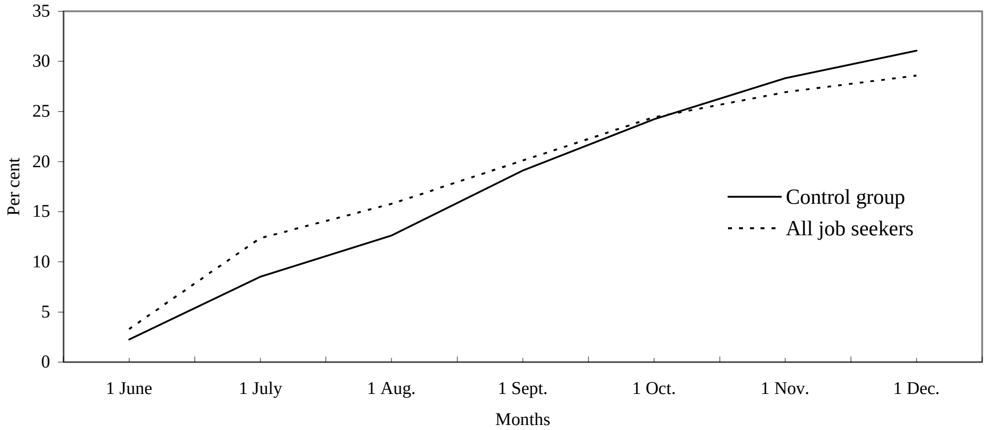
to its matched treated counterpart. Heckman et al. (1997) conclude that in small samples, the choice of matching estimator can make a difference.

In the analysis the experimental control group is replaced by a comparison group drawn from the population of those currently registered at the employment offices on 15 May, and 5 June, 2002. The outcomes for 6,899 individuals represent the counterfactual events in the absence of randomised controls. Previous research has demonstrated the importance of using comparable data in evaluating the results of different studies.[18] Here, information is acquired from the same database for members of the experimental, control and comparison groups. Hence, equivalent characterisation of outcome and explanatory variables is guaranteed.

By comparing the cumulative exits to employment for the control and the comparison groups, we get an idea of to what extent the nonexperimental estimators need to adjust for differences in the outcomes in order to recover the experimental results. According to Figure 1, the exit rates are surprisingly similar considering the above described differences between applicants and nonapplicants. Within six months, 31.1 and 28.6 per cent of the control group and comparison group respectively had achieved employment status. Hence, the tested methods need only adjust for minor deviations in the outcome measure.

[18] See Heckman et al. (1997).

Figure 1. Cumulative exit to employment for members of the control group, and all job seekers



Comparing Experimental and Nonexperimental Results

The first column of Table 6 repeats the regression-adjusted six-month intent-to-treat programme impact presented in Table 4. The results in the three following columns (2-4) refer to different nonexperimental estimates based on outcome differences between a randomly assigned experimental group and a nonrandomly constructed comparison group. If the offered-to-participate indicator is used to impose common support, the nonexperimental Probit estimator (-0.013) deviates by 1.3 percentage points only.[19] This is clearly within the sampling error interval.[20] Note also the higher precision in the non-experimental estimate due to the larger number of observations.

[19] The common-support procedure excludes 626 members of the comparison group, and three members of the experimental group. Since only three experimental group members failed to find a comparable, violating the common-support condition would only have a negligible impact on the estimated programme effect.

[20] A more direct approach in evaluating the bias in nonexperimental estimators, which is applied by Heckman et al. (1997), Heckman et al. (1998) and Smith & Todd (2000), uses data on comparison group members and randomised-out controls. Although similar to the experiment group in observed and unob-

Due to the slightness of the discrepancy, it is not likely that the sample selection model corrects the programme effect for any selection on unobservables. Using the exclusion restriction, "education in desired profession", estimation identifies a positive correction term (Rho). The programme effect is thus adjusted downwards. [21] Since Rho is insignificant, however, this suggests identification without a selection equation.

Columns 3 and 4 report the results of matching experimental and comparison group members on the offer-to-participate probability.[22] The nearest-neighbour estimator produces a point estimate somewhat further from the experimental result. The impact estimate, -0.056, is slightly downward-biased.[23] However, the result is somewhat sensitive to the set of conditioning variables in $P(X)$.[24] Also, performing separate matching on the length of the ongoing registration spell in months, the estimator yields a point estimate (-0.023) considerably closer to the adjusted experimental result.[25]

Note that despite the fact that the nonexperimental Probit and the matching estimates are both based on the same set of covariates, the matching estimate reports a

---

served characteristics, the latter group did not receive any programme services. Hence, a correctly specified nonexperimental estimator should identify a zero programme impact. Performing a probit analysis replacing experiment-group members by randomised-out controls produces a programme effect equal to –0.027 (0.026). The result confirms that the nonexperimental model specification is effective in estimating the true programme impact.

[21] The p-values are 0.012 and 0.487 for the variable "education in desired profession" in the selection and outcome equation respectively.

[22] Both matching estimators perform separate matchings on the relevant start date. Hence, each matched pair started either on 15 May or 5 June. In this way, we ensure similar length of exposure.

[23] Combining data on the randomised-out control group and the nonexperimental comparison group, the estimator generates an insignificant programme effect of -0.059 (0.043).

[24] In contrast, the result is not sensitive to the defined caliper distance applied in the matching.

[25] Heckman et al. (1998) emphasise the benefits of access to information about recent labour force history in the performance of nonexperimental evaluation.

much lower programme impact. This is due to the different weighting schemes of the underlying programme effects. The matching estimator, like the experimental mean-difference estimator, places weight in proportion to the probability of being in the experiment group. Hence, the people most likely to apply for participation are those who get most weight in the programme impact. The Probit estimate, on the other hand, puts most weight in the middle of the probability distribution.[26]

Finally, the result from the kernel-based matching estimator, -0.040, differs by 2.7 percentage points.[27] The benefit from using several comparison group members (instead of just one) is very likely due to the small number of experimental observations. Placing less weight on each particular comparison outcome helps to reduce the uncertainty in the estimated programme effect.

To summarise, the experimental programme results are robust when testing various nonexperimental estimators. The predicted programme effects estimated in a simple Probit model, and two matching procedures, are all fairly close to, and within sampling variance from, the experimental impact estimate. The findings suggest that the available data successfully identifies and adjusts for non-random selection. However, with such

[26] Angrist & Krueger (1999) discuss the weighting issue in depth.

[27] Modifying the applied default bandwidth (0.6) does not dramatically alter the results.

poor precision in the experimental impact estimate, only very large deviations would have generated another conclusion.

Table 6. Parameter estimates on exit-to-employment probability within six months after enrolment, for those *offered* treatment (estimated standard errors in parentheses)

	Experimental, Probit (adjusted) (1)	Nonexperimental, Probit (2)	Nonexperimental, bivariate Probit (3)	Matching[a] (4)	
	Coefficient *Std. error*	*Coefficient* *Std. error*	*Coefficient* *Std. error*	Nearest neighbour	Kernel matching
Programme effect[b]	**-0.013 (0.039)**	**-0.026 (0.024)**	**-0.144 (0.183)**	**-0.056 (0.039)**	**-0.040 (0.029)**
Rho			0.222 (0.245)		
Gender					
Female	-	-	-		
Male	0.039 (0.121)	0.057 (0.039)	0.060 (0.039)		
Age					
18-24	-	-	-		
25-34	0.064 (0.168)	-0.134 (0.055)**	-0.151 (0.058)***		
35-44	-0.088 (0.200)	-0.184 (0.060)***	-0.206 (0.065)***		
45-55	-0.398 (0.234)*	-0.217 (0.066)***	-0.243 (0.072)***		
55-	-0.296 (0.336)	-0.545 (0.080)***	-0.583 (0.090)***		
Educational level					
<Compulsory school	-	-	-		
Compulsory school	-0.340 (0.275)	0.016 (0.076)	0.018 (0.076)		
Upper secondary	0.056 (0.246)	0.156 (0.069)**	0.154 (0.069)**		
University	-0.097 (0.266)	0.175 (0.076)**	0.188 (0.077)**		
Home county					
Big city[c]	-	-	-		

Local labour markets[d]	0.132 (0.177)	0.124 (0.045)***	0.104 (0.051)*
Other	0.030 (0.131)	0.075 (0.041)*	0.063 (0.043)
Experience in desired profession			
No	-	-	-
Yes	-0.061 (0.135)	-0.023 (0.043)	-0.009 (0.046)
Employment type in desired profession			
Part-time, part-time/fulltime	-	-	-
Fulltime	-0.065 (0.112)	-0.024 (0.035)	-0.021 (0.035)
Desired profession			
No classified profession	-	-	-
Management work	1.035 (0.577)*	0.011 (0.165)	0.048 (0.170)
Special theoretical competence	-0.050 (0.276)	0.086 (0.084)	0.129 (0.097)
Short university education	-0.210 (0.276)	0.133 (0.078)*	0.161 (0.084)*
Administrative work	-0.102 (0.277)	0.065 (0.073)	0.085 (0.077)
Service, health care and commercial	-0.137 (0.257)	0.101 (0.063)	0.116 (0.066)*
Craftsman's work	0.051 (0.294)	0.224 (0.075)***	0.235 (0.076)***
Machine work, transport and commun.	0.043 (0.291)	0.120 (0.075)	0.131 (0.076)*
No vocational training required	-0.159 (0.291)	0.101 (0.075)	0.112 (0.076)
UI-compensation			
No	-	-	-
Base premium	0.185 (0.189)	0.288 (0.071)***	0.290 (0.071)***
Income related	0.662 (0.168)***	0.340 (0.055)***	0.330 (0.057)***
Working disability			
No	-	-	-
Yes	-0.294 (0.207)	-0.660 (0.062)***	-0.665 (0.062)**
Citizenship			
Non-Swedish	-	-	-
Swedish	0.078 (0.164)	0.168 (0.061)***	0.151 (0.064)***
Expanded search area[e]			
No	-	-	-
Yes	0.259 (0.142)	0.148 (0.050)***	0.158 (0.051)***

Unemployment history					
No. of LMPs	-0.004 (0.024)	-0.011 (0.008)	-0.009 (0.008)		
No. of periods openly unemployed	0.021 (0.016)	0.018 (0.005)***	0.020 (0.006)***		
Unemployment duration					
Ongoing unemployment period	-0.029 (0.042)	-0.069 (0.011)***	-0.071 (0.011)***		
All unemployment periods	-0.080 (0.027)***	-0.027 (0.007)***	-0.028 (0.007)***		
Status at the experiment start					
Openly unemployed	-	-	-		
In job	-0.216 (0.187)	-0.166 (0.044)***	-0.192 (0.053)***		
In LMP	-0.151 (0.132)	-0.181 (0.046)***	-0.192 (0.047)***		
Start date					
15/5-02	-	-	-		
5/6-02	-0.146 (0.115)	-0.218 (0.034)***	-0.217 (0.034)***		
Constant	-0.399 (0.350)	-0.620 (0.104)***	-0.577 (0.115)***		
Log-likelihood	-357.138	-3660.481	-4782.367		
Pseudo R^2	0.094	0.097			
Number of observations	627	6613	6613	646	7242

Notes: col. (1) refers to regression (Probit) adjusted experimental estimates. col. (2) Probit estimates, nonexperimental. col. (3) bivariate Probit-estimates, nonexperimental. *, **, *** refers to significance at 10, 5 and 1 per cent level respectively. [a]: Standard errors are bootstrapped. Caliper distance used in the nearest-neighbour matching: 0.01. Applied bandwidth in the kernel-based matching: 0.6. [b]: Estimated treatment effects and standard errors in percentage points. [c]: Refers to the counties of Stockholm, Västra Götaland and Skåne. [d]: Refers to the counties of Värmland, Dalarna, Gävleborg, Jämtland, Västernorrland, Västerbotten and Norrbotten. [e]: During the first 100 days of unemployment, a job seeker is allowed to restrict the search area geographically.

Conclusions

This paper has used data from a randomised experiment to investigate the effect of offering job-search activities on the Internet. Experiments on labour market topics are very rare in Europe in general, and in Sweden in particular. Thus, in conducting the first such experiment in Sweden since 1975, the first pertinent question concerns whether or not the evaluation design succeeded in deriving interesting and relevant policy parameters, i.e., did the experiment work or not? In so far as experiment-related problems such as ethical concerns, bureaucratic resistance and randomisation bias were circumvented, the answer is positive. Unfortunately, with the assessed programme services being nonmandatory, a considerable fraction of the applicants either failed to show up or dropped out early in the process. This reduced the value of the experiment. The somewhat unusual selection mechanism also makes it difficult to extract lessons that could be of use in future experiments. One important lesson is, however, to be careful in relying on job seekers´ self-reported interest in participating in an activity, at least when participation is voluntary. In this experiment, as much as 47 per cent of those offered the services never entered the programme, despite the fact that they submitted applications only within a few weeks before. Also, it is quite possible that the compliance ratio would have benefited from an outreach procedure before programme start, in which the specified participation prerequisites could have been confirmed.

What did we then learn about the future of conducting job-search activities on the Internet? Well, with less than a thousand job seekers showing interest in the programme, and only 636 relevant applications submitted, it is likely that the Swedish Labour Market Board misjudged the clients´ interest in these services. Since one of the purposes of the demonstration was to explore the demand for this type of services, this was useful information. However, the small sized sample produced impact estimates with low precision. Also, with the severe compliance problems reducing the expected response differences between the experimental and control group, the insignificant impact estimate was not surpringsly. The results from the nonexperimental dose-response calculations on two different treatment-dose indicators also fail to reject the hypothesis of a zero programme impact. There are, however, some signs of small positive short-term effects for some level of usage of the services. Further investigations would therefore benefit from a more precise targeting of the services. Among those showing interest in the services, and applying for participation, youth, highly educated, and people living in one of the big city areas were overrepresented.

The final consideration concerns the methodological findings of this paper and is connected with the opportunity for assessing nonexperimental estimators. In evaluating the experimental group outcome against the outcome from a constructed comparison group, using various techniques to offset systematic differences, we found standard econometric methods to be successful in reproducing the experimental impact estimate. However, since the estimators needed to adjust for only minor differences in the outcome measure, the results may not be relevant to other programmes involving other (stronger) selection processes. Also, with the imprecise experimental and nonexperimental estimates, only very large deviations would generate the alternative conclusion.

References

Angrist, J. D. and Krueger, A. B. (1999), “Empirical Strategies in Labor Economics”, Ashenfelter, O. C. and Card, D. (eds), Handbook of Labor Economics, Vol 3A, Elsevier, Amsterdam, The Netherlands, pp. 1277-1366

Björklund, A. and Regnér, H. (1996), “Experimental Evaluation of European Labour Market Policy”, Schmid, G. and O’Reilly, J. and Schömann, K. (eds), International Handbook of Labour Market Policy and Evaluation, Edward Elgar, Cheltenham, U.K, pp. 89-113

Bratberg, E. and Grasdal A. and Risa A.E. (2002), “Evaluating Social Policy by Experimental and Nonexperimental Methods”, Scandinavian Journal of Economics, Vol 104, No 1, pp. 147-171

Calmfors, L., Forslund, A. and Hemström, M. (2001), ”Does active labour market policy work? Lessons from the Swedish experiences”, Swedish Economic Policy Review, vol 8, No 2, pp. 61-124

Dehejia, R. and Wahba S. (2002) “Propensity Score Matching Methods for Nonexperimental Causal Studies”, Review of Economics and Statistics, Vol 84, pp. 151-161

Dehejia, R. and Wahba S. (1999) “Causal Effects in Nonexperimental Studies: Reevaluation of the Evaluation of Training Programs”, Journal of the American Statistical Association, Vol 94, No 448, pp. 1053-1062

Delander, L. (1978) “Studier kring den arbetsförmedlande verksamheten”, Arbetsmarknadspolitik i förändring, SOU 1978:60.

Fraker, T. and Maynard R. (1987) “The Adequacy of Comparison Group Designs for Evaluations of Employment-related Programs”, Journal of Human Resources, Vol 22, pp.194-227

Friedlander, D. and Robins P. K. (1995) “Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods”, American Economic Review, Vol 85, pp. 923-937

Heckman, J. J., Ichimura, H., Smith, J. and Todd, P. (1998) "Characterizing Selection Bias Using Experimental Data", Econometrica, Vol 66, No 5, pp. 1017-1098

Heckman, J. J., Ichimura, H. and Todd, P. (1997) "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", Review of Economic Studies, Vol 64, No 4, pp. 605-654

Heckman, J. J. and Smith, J. (1995) "Assessing the Case for Social Experiments", Journal of Economic Perspectives, Vol 9, No 2, pp. 85-110

Heckman, J. J., LaLonde, R. J. and Smith, J. (1999), "The Economics and Econometrics of Active Labor Market Programs", Ashenfelter, O. C. and Card, D. (eds), Handbook of Labor Economics, Vol 3A, Elsevier, Amsterdam, The Netherlands, pp. 1865-2097

Kuhn, P. and Skuterud, M. (2004) "Internet Job Search and Unemployment Durations", American Economic Review, Vol 94, No 1, pp. 218-232(15)

LaLonde, R. (1986) "Evaluating the Economic Evaluations of Training Programs with Experimental Data", American Economic Review, Vol 76, No 4, pp. 604-620

Martin, J. P. and Grubb, M. (2001), "What works and for whom: a review of OECD countries´ experiences with active labour market policies", Swedish Economic Policy Review, vol 8, No 2, pp. 9-56

Raaum, O., Torp, H. and Goldstein, H. (1994) "Experiments in Manpower Evaluation: The Case for Simple Estimators? Experience from a Norwegian Study of Labour Market Training", Sosialekonomisk Institutt, Oslo University

Rosenbaum, P and Rubin, D. (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects", Biometrika, Vol 70, No 1, pp. 41-55

Rosholm, M. and Skipper L. (2003) "Is Labour Market Training a Curse for the Unemployed? – Evidence from a Social Experiment", IZA Discussion Paper, No 716

Smith, J. and Todd, P. (2004) “Does Matching Overcome LaLonde’s Critique of Non-experimental Estimators?” Journal of Econometrics, No 125(1-2), pp. 305-353

Torp, H., Raaum, O., Hernaes, E. and Goldstein, H. (1993) “The first norwegian experiment”, Jensen, K. And Madsen, P. K. (eds), “Measuring Labour Market Measures”, Danish Ministry of Labour, pp. 97-140

Van Den Berg, G. J. (1994), “The Effects of Changes of the Job Offer Arrival Rate on the Duration of Unemployment”, Journal of Labor Economics, No 12, pp. 478-498

Van Den Berg, G. J. and van Der Klaauw, B. (2001), “Counselling and monotoring of unemployed workers: Theory and evidence from a controlled social experiment”, IZA Discussion Paper, No 374

White, M. And Lakey, J. (1992) “The restart effect: Does active labour market policy reduce unemployment?”, Policy Studies Institute, London