Targeting Labour Market Programmes – Results from a Randomized Experiment

Stefanie Behncke[a], Markus Frölich[b] and Michael Lechner[c d]

1. Introduction

Profiling and Targeting Systems have received considerable attention in recent years, both from academic researchers as well as from policy makers. These are statistical systems suggesting who should receive certain public services, who should be offered re-employment bonuses (O'Leary, Decker, and Wandner, 2005), who should attend certain active labour market programmes,[1] who should be searched at the airport to maximize Airport Security (Persico and Todd, 2005; Manski, 2006), or which treatment (punishment) should be given to

a   Stefanie Behncke, Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Varnbüelstr. 14, CH-9000 St. Gallen, Switzerland, Stefanie.Behncke@unisg.ch
b   Markus Frölich, Chair of Econometrics, University of Mannheim, D-68131 Mannheim, Germany, Markus.Froelich@uni-mannheim.de. Markus Frölich is also affiliated with IZA, Bonn, SEW, St. Gallen, and IFAU, Uppsala.
c   Michael Lechner, Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Varnbüelstr. 14, CH-9000 St. Gallen, Switzerland, Michael.Lechner@unisg.ch, www.sew.unisg.ch/lechner. Michael Lechner has further affiliations with ZEW, Mannheim, CEPR and PSI, London, IZA, Bonn, and IAB, Nuremberg.
1   See the book "Targeting Employment Services" (Eberts, O'Leary and Wandner, 2002) and the articles therein, Berger, Black, and Smith (2001), Black, Smith, Plesca, and Plourde (2002), Bryson and Kasparova (2003), Dehejia (2005), Eberts (2002), Fraser (2000), Keum (2001), Manski (2000, 2004, 2007), Moisala, Suoniemi, and Uusitalo (2006), Plesca and Smith (2005), Rudolph and Müntnich (2001), Staghøj, Svarer, and Rosholm (2007), and Stephan, Rässler, and Schewe (2006).

certain criminal offenders, to name just a few examples.[2] Particularly in the area of provision of public services such systems are considered as a potential means to target services more directly to clients in need or to those who would benefit most from it. The increasing use of such profiling and targeting systems is made possible through the widespread availability of PCs connected to the Intra- or Internet in most government offices. These make the online provision of individual impact predictions possible.

The interest in the targeting of active labour market programmes (ALMP) has been triggered by a number of previous disappointing evaluation results. ALMP have been introduced in many countries during the 1990s to combat the problems of high and persistent unemployment or low earnings of disadvantaged groups. The initial enthusiasm for this paradigm waned when many evaluation studies (in various European countries) concluded with finding rather moderate or even negative treatment effects. This emphasized the need for targeting ALMP to those unemployed persons who may actually benefit from them. Such profiling and targeting systems are, or have been, in use in Australia, South Korea, the Netherlands, and the USA. Several other countries like the UK, Germany, Denmark, Finland, and Sweden are currently piloting such systems or consider their use.[3] Caseworkers seem to have been hostile to such systems in every case, though. It is one of the purposes of this paper to add practical knowledge useful for the design of such systems.

Profiling and Targeting systems generally serve two purposes: First, they provide information to the caseworker. Econometric impact estimates of expected programme benefits can be provided on a relatively disaggregated level. For example,

---

2   Further examples from biometrics include the choice among various medical drugs for cancer treatment (or its dosage) or the choice of a rehabilitation therapy for alcohol related problems. Past sickness history and intermediate outcomes are used to adjust a time varying dosage. For references on targeting of treatments in biometrics and related fields, see Brownell and Wadden (1991) on obesity, Velicer and Prochaska (1999) on smoking, Murphy (2003) and Murphy, Lynch, Oslin, McKay and TenHave (2007) on drug and alcohol dependence, and Rush (2005) on depression.

3   Germany is currently piloting a targeting system, see Stephan, Rässler, and Schewe (2006). Sweden is currently piloting a profiling system in one county and is considering a targeting system as well. (Personal communication by Anders Forslund, IFAU, Uppsala.) Finland has piloted a profiling system and is about to implement it. (Personal communication by Roope Uusitalo, Labour Institute for Economic Research, Helsinki; see also Moisala, Suoniemi, and Uusitalo, 2006.) Denmark is considering a targeting system, see also Staghøj, Svarer, and Rosholm (2007). The UK used profiling in a pilot study for workers on incapacity benefits and is currently rolling out the system. (Personal communication by Alex Bryson and Richard Dorsett, Policy Studies Institute, London.)

the system can provide individual estimates, for a particular client i, of the benefit from taking part in labour market programme A instead of B or C. Their second purpose is to solve an agency problem in a decentralized bureaucracy: Caseworkers may pursue their own strategies or aim to satisfy their own beliefs or convictions, which may not be fully aligned with those intended by the law or the central government. For example, the Swiss federal law contains relatively detailed provisions when to use certain sanctions if the unemployed person displays insufficient job search efforts. However, the actual implementation of this sanctioning policy differs substantially between caseworkers and between employment offices (LALIVE, VAN OURS, and ZWEIMÜLLER, 2005; EGGER and LENZ, 2006; FRÖLICH et al., 2007). Similarly, regarding the choice of active labour market programmes, caseworkers are often much less concerned about programme costs than taxpayers are, and they may have different aims than intended by the law, e.g. place more emphasis on sustainable reintegration instead of rapid reintegration.

Profiling and Targeting Systems can be implemented in several ways, with the amount of discretion left to the caseworkers being a crucial parameter. In the one extreme, caseworkers have no discretion and the statistical system determines which actions are to be taken for a particular client. In the other extreme, the statistical system may simply act as an information tool leaving the choice entirely at the caseworker's discretion. Restricting caseworkers' discretion, as done for example in the Worker Profiling and Reemployment Services Initiative (WPRS) in the USA (EBERTS, O'LEARY and WANDNER, 2002), has the advantage of streamlining service provision and ensuring equal treatment of clients across sites. It will lead to an alignment of procedures and actions across offices and caseworkers and thereby reduce the heterogeneity in the way the law is implemented. It may also save on caseworker's time, e.g. when clients try to "negotiate" with their caseworker to receive or avoid certain services or measures. On the other hand, it may severely reduce job satisfaction and curtail intrinsic motivation of caseworkers who see themselves as subordinates of a computer. This point strongly supports a high degree of caseworker discretion, where the statistical system acts to *assist* the caseworkers by providing additional information and suggestions for possible actions. Another important advantage of caseworker discretion is its flexibility in permitting private information of caseworkers to enter the choice of services. Caseworkers may have obtained detailed observations of a client's motivation, personality, work ethos, which could in principle also be made available to the statistical system but presumably only with considerable measurement error. If such private information is available, it is desirable that caseworkers can deviate from the recommendations of the statistical system. In this paper, we find evidence that an implementation with full caseworker discretion is not likely to work

well. Without strong incentives or coercion, caseworkers may ignore or sabotage a system that is intended to provide information only.

A large randomized field experiment was conducted in 2005 in several employment offices in Switzerland, where part of the caseworkers had access to a statistical system providing individual predictions of unemployment risk in relation to participation in different programmes of the Swiss active labour market policy. The field study was motivated by previous work of Frölich, Lechner, and Steiger (2003), and Lechner and Smith (2007) who found indications for substantial potential benefits of introducing such a statistical system in Switzerland.[4] During the randomized field study, the behaviour of those caseworkers who had access to the system and their treatment choices were monitored and compared to a randomized control group. The evaluation results show that caseworkers did not change their behaviour in any significant way due to having access to the additional information. Caseworkers either decided to ignore the system or were over-confident in that their own experience clearly dominates any information that a statistical system might provide.

Hence, there is a dilemma if one intends to use profiling and targeting to solve the agency problem in a decentralized bureaucracy: When providing no incentives to use the system, caseworkers may ignore or sabotage it. However, severely restricting caseworkers' discretion crowds out intrinsic motivation and does not exploit the value of the private information of the caseworker. (Even the fear that the system could potentially be used to restrict caseworker discretion could already raise strong objections on the side of the caseworkers, as was the case with SOMS in Canada, which had to be stopped even before it was fully operational, Colpitts, 2002.) Our findings clearly suggest that permitting full caseworker discretion is unlikely to work, at least when piloting the project. In order to evaluate the impact of the statistical system, incentives for the caseworkers should be set to comply the predictions.

In the next section, we describe profiling and targeting systems for unemployed and some of their implementations. Section 3 explains the particular implementation of the Swiss targeting system. Section 4 gives the detailed results of the experimental evaluation of the Swiss system and Section 5 concludes. Four appendices provide further details concerning the data, the econometrics used for the predictions, the Swiss active labour market policy, and the results of the experimental evaluation.

---

4   Both studies found indications for treatment effect heterogeneity regarding employment chances that caseworkers did not exploit. Caseworkers did not appear to be more successful in selecting labour market programmes than a purely random allocation.

2. Profiling and Targeting Systems

2.1 Optimal Programme Choice

In a series of recent papers, Manski (2000, 2004, 2007) considered the choice between different treatments from the perspective of a social planner. A number of mutually exclusive treatments is available and the social planner attempts to choose the optimal treatment for each client.[5] The treatments may be different medical drugs, different therapies for persons with alcohol-related problems, or different active labour market programmes (ALMP) to mention just a few examples. The ALMP available often comprise job search training, personality courses, language courses, computer courses, vocational skills training, further training, employment programmes etc.

At some time t the individual i may receive one of $R+1$ different treatment options and we observe an outcome at the time (or during the period until) $t+\tau$. Let

$$Y_{i,t+\tau}^0, Y_{i,t+\tau}^1, \ldots, Y_{i,t+\tau}^R$$

be the *potential* outcomes for individual i, i.e. those outcomes that would be observed if a particular treatment is chosen. The treatment 0 usually refers to the choice of "no active treatment", e.g. no medical drug or no labour market programme. (In our application, we consider repeated treatment choices, where a choice is made at every counselling meeting. Treatment 0 then means "no programme is chosen now" but perhaps at the next counselling meeting.)

If the treatment choice cannot, or should not, be delegated to the individual, e.g. because of the moral hazard problem inherent in the unemployment insurance system, the social planner can only allocate programmes on the basis of observed characteristics, which we denote by X_{it}. Under certain conditions (absence of risk aversion on the planner's side or a binary outcome variable,[6] absence of supply side constraints, absence of externalities), the optimal choice for a client with characteristics X_{it} is given by

---

5   We examine here only the choice of programmes as one of the instruments of active labour market policy. Wunsch (2007) considers the design of an optimal unemployment insurance system as a whole where also other instruments are included.

6   When the outcome variable is binary, the conditional mean characterizes the entire conditional distribution.

$$r^*(X_{it}) = \arg\max_{r \in \{0,\dots,R\}} E[Y^r \mid X = X_{it}]$$

and is thus based on estimates of the expected conditional potential outcomes

$$E[Y^0_{i,t+\tau} \mid X_{it}], \dots, E[Y^R_{i,t+\tau} \mid X_{it}].$$

This is the basis for statistical *targeting* systems, which select for each client the programme with the highest expected outcome.

This is in contrast to *profiling* systems, which predict only the outcome $E[Y^0_{i,t+\tau} \mid X_{it}]$ i.e. when "no active treatment" is chosen. In the particular application to active labour market programmes, the profiling systems often estimate the *risk of becoming long-term unemployed if not taking part in ALMP*. Unemployed persons assessed to be at high risk are then assigned to the most intensive services. As discussed e.g. in FRÖLICH, LECHNER, and STEIGER (2003), targeting is preferable to profiling if a variety of heterogeneous labour market programmes are offered, as it is often the case in European countries, and if the long-term unemployment risk is not highly correlated with programme impacts, as it was found e.g. in BERGER, BLACK, and SMITH (2001).

In practice, the social planner will not be omnipresent and needs agents, i.e. the caseworkers in the employment offices, to implement the intended policies. In addition to providing information on vacancies, monitoring the unemployed and providing psychological support, they should choose labour market programmes to maximize the objectives of the central planner. This can lead to deviations from the idealized situation described above. First, agents might deviate from their principal's objective. Indeed, when asked in a written questionnaire, around 40% of Swiss caseworkers reported 'improving their clients personal situation', while 35% reported 'long-run reintegration' and 25% 'rapid reintegration' as their primary goal for low qualified unemployed (FRÖLICH et al., 2007). This is at odds to the Swiss unemployment insurance law, which heavily weighs rapid reintegration over long-run reintegration and does not place any weight to the personal wellbeing of the unemployed at all. Second even in the absence of principal-agents problems, they have to estimate the unknown potential employment outcomes for each programme and each particular client. These estimates need to be updated at a regular interval since the characteristics X_{it}, such as the unemployment duration and employability, change over time, which may affect the optimal treatment choice. Caseworkers will usually attempt to predict programme impacts by combining results from past national or regional evaluation studies (which will usually be very aggregated, e.g. for men versus women, young

versus old) with their own observations of the careers of their clients, complemented by discussions with their colleagues and supervisors. For producing these predictions, they face the problem of a relatively small sample size that is available to them, i.e. the number of clients they have personally counselled. This is particularly acute for caseworkers with little job experience. Small sample sizes are even more reduced by a limited capacity of caseworkers to follow-up on their clients after they are de-registered from the employment office. While they usually know their exit state, e.g. to employment or out of labour force, they often do not observe their subsequent career. For instance, if their clients become unemployed again, they might be allocated to another caseworker or be registered in a different employment office, or even worse, might not register at all if they are not entitled to further benefits. However, without observing the subsequent career of the previously unemployed person, the caseworker cannot learn anything on the programme effect.

Indeed, there is evidence suggesting that caseworkers did not succeed in estimating employment probabilities between the various programmes or followed other aims than employment maximisation of their clients: Bell and Orr (2002) found that caseworkers did not systematically select those into treatment who would benefit most from it, and Frölich, Lechner and Steiger (2003) and Lechner and Smith (2007) found similarly that the treatment effect heterogeneity was not successfully taken account of by caseworkers.

A statistical system may thus be helpful to provide estimates of the expected potential outcomes on an individual basis, if a detailed administrative database of past service provision and treatment choices is available. The database will often be nationwide, covering the entire population with detailed individual data and complete follow-up information. An individual who moved to a different locality may be out of sight for the previous caseworker but subsequent unemployment spells will be recorded in the data set available to the statistical system. The large sample size and the complete follow-up information can thus lead to impact estimates that contain *additional information for the caseworkers*. If made available to them, caseworkers may then combine this information with their own expectations of programme impacts, where they can incorporate additional private information about their subjective assessment of the motivation, personality, and work ethos of the unemployed client.

Statistical Profiling and Targeting Systems may often also have another purpose than merely providing information. They may help to streamline service provision and ensure that caseworkers aim at the same target variables as the central planner does. Thereby, the statistical system may help to solve the agency problem in a decentralized bureaucracy. Caseworkers often have considerable

discretion in the actions they take (or ignore to take), e.g. assigning programmes or imposing sanctions. The caseworkers and employment offices may follow their own convictions of what constitutes the best attitude towards unemployed. Some caseworkers pursue rapid re-employment, whereas others grant the unemployed more time to find good job matches. Some caseworkers expect active labour market programmes to be beneficial for immediate employment, whereas others use them also to provide psychological stabilization and develop 'fitness for the labour market' (Egger and Lenz, 2006; Frölich et al., 2007). In addition, a sympathetic or antipathetic relationship to the unemployed person may also unconsciously influence the actions taken. Furthermore, caseworkers probably are less concerned about programme costs than taxpayers would be. The central government provides certain indicators it seeks to pursue but leaves the employment offices almost complete freedom in their implementation (Egger and Lenz, 2006; Frölich et al., 2007). The statistical targeting system, on the other hand, aims at a clearly defined targeting indicator, usually employment (or unemployment) at a certain point in time or over a certain period.

2.2 Experiences with Targeting Systems in Canada and the USA

Canada developed the Service and Outcome Measurement System (SOMS) between 1994 and 1999 as a support system for service delivery staff who retained full discretionary power (Colpitts, 2002). SOMS was based on a set of parametric statistical models of subgroup employment services impacts estimated on a huge database constructed by merging a number of different administrative datasets. SOMS, however, was never implemented mainly because of data security concerns and because of resistance from frontline caseworkers who perceived SOMS as a threat to their own job security. Staff fears were fanned by an impending organizational restructuring within the ministry for human resources and employment. The SOMS database was deleted in 2002, before SOMS ever was tried in the field.

The Worker Profiling and Reemployment Services (WPRS) system was implemented nationwide in the USA in 1994 and has operated in all states since that time. WPRS ranks new UI beneficiaries who are not exempt from UI job search requirements by their probability of UI benefit exhaustion. UI beneficiaries ranked by WPRS are then referred automatically to reemployment services in order from highest to lowest probability until the available slots for services are filled. Most states rank UI beneficiaries exhaustion probabilities using a logit model of benefit exhaustion (Wandner, 2002). Evaluations have produced a range of impact estimates for WPRS. Dickinson, Decker, and Kreutzer

(2002) estimated reductions in UI benefit receipt as large as half a week based on analysis of state administrative data. Black, Smith, Berger, and Noel (2003) estimated a reduction of 2.2 weeks in benefit duration based on random assignment to WPRS in Kentucky.

A Frontline Decision Support System (FDSS) was developed for the state of Georgia in the USA. It was planned as a decision support system for frontline staff in one-stop employment service centres to target reemployment services. The FDSS included two main parts: the systematic job search module, and the service referral module. The service referral module identifies the sequence of activities that most often lead to successful employment for clients with similar background characteristics (Eberts and O'Leary, 2002). It was pilot tested in two Georgia Career Centers but discontinued soon after for several reasons, without subsequent impact evaluation.[7]

The Work First Profiling Pilot Project (WFPPP) involved comparison of targeted assignment with random assignment of referrals to reemployment services among Temporary Assistance for Needy Families (TANF) benefit recipients in Michigan. A net impact evaluation suggested that targeting yielded a cost effective improvement in employment success (Eberts, 2002).

In the following section, the evaluation of a pilot study of a targeting system in Switzerland is described, where caseworkers retained full discretion in their treatment choices.

3. The SAPS Experiment in Switzerland

3.1 Background

Switzerland enjoyed very low unemployment rates during most of the last century until the recession of the early 1990s when unemployment increased to levels not seen before. This triggered a complete revision of the unemployment insurance act in 1996. In concordance with the conventional wisdom of the OECD at that time, Switzerland switched from a passive unemployment benefit system towards an active system promoting training and work experience to unemployed

---

7   FDSS arrived the same time as a huge rise in UI claims associated with a major recession, a change in mainframe computing environment, which bogged down simple administrative tasks for staff, and the transfer from the Georgia agency of the main FDSS champion who was the assistant commissioner for employment services. She left the agency to lead the U.S. Department of Labor southern regional office. We thank Chris O'Leary for pointing out many of these details.

persons. The federal states (cantons) were forced to provide a minimum number of active labour market programme places, and participation was made mandatory for every unemployed person if allocated to a programme by the caseworker. Allocation to a programme is at the caseworker's full discretion, and non-compliance leads to a suspension of benefit payments.

The initial enthusiasm about ALMP has waned in the recent years since several evaluation studies found rather moderate or even negative effects. While some policy actors sympathized with the idea of abolishing the (most) expensive measures, the prevailing view was that active labour market programmes should remain, but should be better targeted towards those who clearly benefit from them.

This laid the foundation for the SAPS experiment. The Swiss State Secretariat for Economic Affairs initiated a pilot study on targeting active labour market services in 21 employment offices: Caseworkers should be assisted in their treatment choices with statistical information.[8] The idea of the Statistically Assisted Programme Selection (SAPS) was to predict for each individual, which programme would benefit him or her most. Statistical predictions about the net impact should be made for every jobseeker and for every possible labour market programme based on an extremely large and rich database. The database is described in appendix A.

A feasibility study was carried out in 2002 for the State Secretariat for Economic Affairs to explore the possibilities of a potential pilot study. The State Secretariat thereafter continued with the preparations for a pilot study, which was carried out in the field from May/June to December 2005. The field study took place in five different regions: Basel, Bern, St.Gallen, Zürich, and Geneva, where the results for Geneva should be considered with caution since a separate profiling tool was tested in that period and since only two very specialized employment offices participated in the SAPS evaluation. (There were also further problems with Geneva as discussed later.)

Since the pilot project was conducted with the aim to explore the potential for a subsequent large-scale introduction of the statistical system in Switzerland, an impact evaluation of SAPS was a key element. A *randomized* evaluation was featured where caseworkers were randomized either into the *in* or the *out* group. A randomization at the level of the caseworker was preferred to a randomization at the level of the employment office or at the level of the unemployed person. The number of participating employment offices was considered too small for a

---

8   A very detailed report (in German) is given in BEHNCKE, FRÖLICH, and LECHNER (2007).

Figure 1: Unemployment Rate January 1990 – August 2007 in Percent

%

6

5

4

3

2

1

0

1990 91 92   93   94   95   96   97   98   99 2000 01   02   03   04   05   06   07   08

reasonable randomization across sites. A randomization at the level of the unemployed person, on the other hand, would have led to the problem that a caseworker would receive employment predictions for some of his clients but not for others. Such a situation would presumably have reduced compliance and/or would also have produced spill over effects: a caseworker receiving predictions for some of his clients would try to extrapolate those predictions to his other clients, for whom no predictions were provided.

Therefore, in each pilot office half of the regular caseworkers were randomized in and half out. A third group of caseworkers, called the *definitive nonparticipants*, contains all the caseworkers who were not included in the randomization for various reasons.[9] This should also imply a randomization on the unit of the jobseekers because at a given point in time, each jobseeker is assigned uniquely to one caseworker. Randomization, however, is only implied at the beginning of the field study since newly registered jobseekers could have been assigned by the

---

9   Before the randomization was carried out, a number of caseworkers were removed to restrict randomization only to the 'regular' caseworkers. Those persons include the management of the employment office, all caseworkers with substantive administrative duties, caseworkers that mainly assist only specific groups of unemployed (e.g. disabled, youth, unemployed with the intention of self-employment), caseworkers that were on sick leave for a longer time or about to retire in the next months or were known to leave the employment office soon, and caseworkers, who announced their unwillingness to participate in the field study *before* the randomization took place.

office management in a non-random way to treatment and control group. We therefore distinguish in the later analyses between a *stock* and a *flow* sample. The stock sample contains all jobseekers who were already registered at the beginning of the field study in the respective employment office. The flow sample contains all jobseekers who entered later. Whereas the stock sample should (almost[10]) be randomized, the assignment process for the flow sample was beyond our control. The employment office managers could have changed the allocation between caseworkers and new jobseekers such that the more difficult cases were assigned to the treatment group (which had access to the tool) or rather to the control group (which did not experience the burden on their time resources by participating in the field study). The ensuing groups might thus in principle not be randomized. (The equality of means tests, however, shows that such concerns are not justified.)

Apart from concerns about randomization, there is also an independent interest to evaluate the impact of SAPS predictions for the stock and the flow separately, because they become available at different times in the unemployment spell. For the flow sample, the predictions are made right from the beginning of the spell, whereas they start for the stock sample at the beginning of the experiment, i.e. after they have been unemployed for a while. From this perspective, the analysis of the flow sample would be more interesting since a practical implementation of a SAPS system should provide predictions as early as possible. Yet, any differences in estimated impacts could also be attributable to declining or increasing interest of the caseworkers during the course of the field study.

Table 1 shows the number of caseworkers in the pilot offices. The first two columns give the number of caseworkers randomized in and out. For some of the caseworkers it turned out that they actually did not counsel any unemployed during the period of the field study, e.g. due to sickness, leave, re-organization, or complete focus on employer contacts. The numbers in brackets provide the number of caseworkers with non-zero clients during the field study. (There is still quite some variation between them, with some caseworkers counselling only 10 to 20 clients at a given time, whereas the upper limit is about 150.) The last two columns show the number of official dropouts of caseworkers during the period of the field study. These are caseworkers whose number of clients officially falls to zero due to retirement, prolonged sickness, dismissal, quit, or death. (It does not include caseworkers who were no longer interested in participation.)

---

10   There was a time delay of a few weeks between randomization of caseworkers and beginning of the pilot study.

Table 1: Randomization of Caseworkers in the Pilot Offices

	Randomization		Dropout during study	
	In	Out	In	Out
Region Basel: Basel city (3 offices together)	29	29	1	1
Region Bern:				
Jura bernois and Bienne[a]	16 (15)	13 (11)		2
Bern Centre	9	8	1	1
Bern West	12	11	1	2
Bern Gümlingen	8	8		
Bern Zollikofen	8	7		
Bern Betlehem	11 (8)	10 (8)	1	
Tafers (Fribourg)	4	3		
Murten (Fribourg)	3	3		
Region Zürich:				
Lagerstrasse	7	6		1
Bülach	9	9	1	1
Region St.Gallen:				
Oberuzwil	9	7	1	3
Sargans	4	4	2	
Region Geneva:[b]				
Gavard	5	7		2
Rive	12	12 (11)	2	2
Total:	146 (142)	137 (132)	10	15

Note: The first two columns give the number of caseworkers randomized in and out. For some of the caseworkers it turned out that they did not counsel any unemployed for various reasons. The numbers in brackets provide the number of caseworkers with non-zero clients during the field study.

The last two columns show dropout of caseworkers during the study due to retirement, prolonged sickness, dismissal, quit, or death.

a) These are 2 offices plus 3 sub-branches. Randomization was carried out together because some caseworkers were employed at the same time in two offices or branches.

b) The employment offices in the canton Geneva are not geographically organized but specialize on occupation and skill groups. The two employment offices Gavard and Rive were specialized on specific occupational groups.

These fluctuations are more pronounced in the two employment offices of St. Gallen where 25% of the caseworkers left the office during the period of the field study.

3.2 The SAPS Estimates and the Outcome Variable

During the field study, the participating caseworkers received suggestions for the *optimal* programme for every particular jobseeker. The SAPS predictions were made available to the caseworkers via an easy-to-use Internet application. Based on a personalized login, caseworkers had access only to the predictions of their own clients. They were asked to provide feedback about these predictions. (Every access to the database was recorded.) The caseworkers of the control group had no access to the predictions. For the purpose of later analysis, we also computed the employment predictions for their clients, but made them not available.

The predictions were updated every second week by incorporating new information on time varying covariates (in particular unemployment duration). This is a big advantage vis-à-vis simple profiling models as it takes into account that the optimal time *when* a labour market programme should start may also vary across individuals.

The overriding aim of active labour market policies in Switzerland is rapid, and ideally sustained, employment. The outcome variable of most interest was therefore defined as the *number of months in stable employment in the next 12 months*, where an employment spell is considered as 'stable' if it lasts for at least 3 months. This outcome variable was used for the predictions given by SAPS. For reasons of consistency, it is used for the evaluation of effects of SAPS as well.

The active labour market programmes were grouped into 6 to 8 categories (depending on region). See also Appendix B and C. For every individual i with characteristics X_{it} every second week it was predicted how many months of stable employment would be expected if that individual started a programme of this category now. A further category was the option 'no programme now, but perhaps later'.

The information conveyed by SAPS to the caseworker consisted of two parts: First, for every programme the expected number of months in stable employment was predicted. Second, the statistical precision of the estimates was also conveyed to the caseworker in that the set of all programmes was divided into three groups: The significantly best programmes, the intermediate programmes, and the worst programmes. The *set of significantly best programmes* contains the true programme with a relatively high statistical confidence. This set was estimated by Multiple

Comparison with the Best (MCB) routines (see Horrace and Schmidt, 2000; or Frölich, 2008).

The cardinality of the set of significantly best programmes varies across jobseekers: For some jobseekers, there was only one programme being statistically significantly better than the rest. For other jobseekers, this set contained several programmes or, in some cases, all programmes. It was suggested to the caseworkers to choose an option from the *set* of significantly best programmes with a slight preference to be given to the programme with the largest estimate. (This distinction will be used later when we consider the compliance of the caseworkers.) Important aspects on the econometric methodology are given in Appendix B, with more information to be found in Frölich (2008) and the corresponding (more detailed) discussion paper Frölich (2006).

4. Evaluation of the Experiment

4.1 The Field Study and the Presumed Causal Chain

The field study took place from May/June to December 2005. The evaluation of the impact of SAPS is based on data from the unemployment insurance system until December 2006. In total, employment predictions were made for 18 713 jobseekers whereas the control pool contained 16 677 jobseekers in pilot employment offices during this period. In a first instance, we examined whether the control and treated jobseekers are similar in their observed characteristics. Therefore, we examined separately the *stock* (= 22 758 jobseekers) and the *flow* sample (= 12 632 jobseekers).[11] All offices passed the randomization test with the exception of the two offices in Geneva. (See Behncke, Frölich, and Lechner 2007 for more details.) It seems that a re-allocation of caseworkers had taken place in Geneva after our randomization. Therefore, Geneva is treated separately in the following. As we will see later, the offices in Geneva also had a very low compliance rate, such that any econometric analysis for these caseworkers would not have been sensible anyhow.

Apart from the distinction between *stock* and *flow* sample we also have to consider the possibility that a jobseeker's caseworker may change over time. Apart from random fluctuations e.g. due to extended holidays, sickness, quit, dismissal,

---

11 As described in the previous section we randomized the *caseworkers* in each office, not the jobseekers themselves. Hence, randomization at the jobseeker level was not fully under our control.

or retirement of the caseworker, the policy of case-worker-change (Dossierwechsel) is a major reason of this. In many offices, the caseworker is changed if the jobseeker stays unemployed for more than 6 or 8 months, with the intention to introduce new ideas in the job search process. This implies that those jobseekers for whom a caseworker change is observed are more likely to have difficulties in finding a job, which is clearly seen in their unemployment histories. In most of the analyses, we will define treatment status of the jobseeker as *time-invariant* according to the randomization status of the *first* caseworker. More precisely, for the stock sample it is the caseworker in charge at the inception of the field study, whereas for the flow sample it is the first caseworker observed after the beginning of the study. The time-invariant definition will implement the intended randomization design as closely as possible. However, it also implies that some of the controls may actually have later on been influenced by SAPS predictions, whereas for some of the treated, SAPS predictions were no longer available. In total, there were 2263 treatment-group switchers. In addition to the total sample, we will therefore also consider the subsample *without switchers*, i.e. after deleting all jobseekers for whom a change in caseworker between treatment and control group has been observed.[12] For the remaining subsample the SAPS predictions were either available during the entire field study or never at all. Although the switchers are clearly different in their observed characteristics from the non-switchers, we do not find any evidence for systematic differences between those who switch into treatment versus those who switch into the control group. See also Appendix D.1.

In evaluating the field study, we follow the hypothesized causal chain of the treatment. Caseworkers received a one-hour introduction into the use of the Internet SAPS system, which in itself is unlikely to have affected their counselling style. Thereafter caseworkers were free to download the employment predictions for any client at any time, which was recorded by our database. Caseworkers were also asked to provide feedback online, which was also recorded.

The hypothesized causal chain, which is sketched in the following figure, thus starts with the download of the SAPS predictions for a particular client.[13] Caseworkers usually downloaded them during or before a counselling meeting. The download of the predictions could have had two effects: They could have affected the counselling style of the caseworker, who e.g. might have shown these

---

12 We do not exclude jobseekers with caseworker changes where both caseworkers belonged to the treatment group or both to the control group.

13 There may also have been a spill-over effect from treated to control caseworkers, for which we do not observe any anecdotal evidence.

predictions directly to the unemployed person. Second, they could have influenced the choice of ALMP, perhaps not immediately but within the next few weeks. From the overall feedback that we received from the caseworkers, it seems that their counselling style has been only very little affected, if at all. Hence, the second channel, i.e. the choice of ALMP, is the channel that we consider most relevant here.

Download of the SAPS predictions   →   Choice of ALMP

↓   ↓

Counselling style   →   Employment status after one year
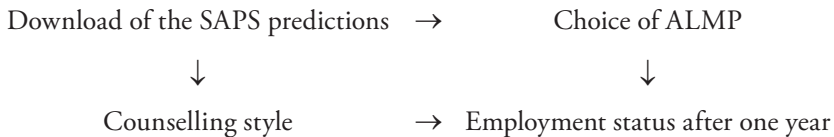
Appendix D.2 provides some details on the frequency of downloading SAPS predictions. Overall, the SAPS predictions were examined at least once for 37% of all jobseekers in the treatment group. Hence, for almost two thirds of the jobseekers the predictions were never viewed and these jobseekers could thus hardly have been influenced by the pilot study. In Geneva, the SAPS project was almost completely ignored: the SAPS predictions were hardly ever examined at all. Therefore, Geneva is omitted in most of the following analyses.

4.2 The Choice of ALMP between Treated and Control

Since the main impact of SAPS on employment is expected by changing the allocation of jobseekers to ALMP, we start with examining the choice of programmes between the treated and the control jobseekers. In a second step, in Section 4.3 we will also take into account that the SAPS predictions could only have had an effect on those jobseekers for whom predictions had been downloaded.

One would expect that caseworkers in the treatment group would (at least partly) follow the predictions after they have become available to them. Whether a caseworker followed or complied with the predictions can be defined in various different ways. The simplest definition considers a SAPS prediction to have been followed or complied with if the recommended labour market programme according to SAPS has subsequently also been assigned to the jobseeker. In fact, below we consider a number of different definitions of *compliance*, which all have in common that we examine, for each jobseeker, whether SAPS predictions and actual assignment to labour market programmes coincided. We will see that all these different definitions lead to the same conclusion: compliance with SAPS was low or inexistent.

For the following tables it is important to note that although the caseworkers in the control group had no access to the SAPS predictions, we nevertheless computed these predictions also for the jobseekers of the control group. Thereby we can measure "compliance" also for the control group, which is the probability that the hypothetical predictions for the control group coincide with the actual choices made for them, without knowing the predictions. Hence, we can examine whether caseworkers of the treatment group were more likely to follow the SAPS predictions than those of the control group.

Here, we consider the *short-term compliance* with the SAPS predictions. In Definition 1, compliance is defined for a jobseeker *i* if the most recommended programme of the first SAPS-prediction is identical to the first labour market programme assigned within the following 90 days. If no ALMP is assigned within these 90 days, this is considered as "no programme" having been assigned. In Definition 2, compliance is defined for a jobseeker *i* if *any* of the set of MCB best programmes is identical to the first labour market programme assigned within the following 90 days. Definition 2 thus subsumes Definition 1.

In principle, it is possible that the 90 days window is too short and caseworkers needed more time to implement the SAPS recommendations. Similarly, they might have complied with the SAPS predictions but only with the second ALMP they assigned. For example, they might have had another ALMP in mind for that client and sent him to that programme first before implementing the SAPS recommendations with a second programme. To permit for such delayed compliance, we will also consider alternative definitions in Appendix D.3, which we refer to as long-term compliance.

Table 2 gives the respective short-term compliance rates for treated and controls. The assigned ALMP coincided with the programme with the highest SAPS prediction for only 12% of the jobseekers. If the wider Definition 2 is used, this rate increases to 29%. Strikingly, these rates are the same for the treatment and the control group. Since the control group reflects what would have happened in the treatment group if the SAPS predictions were not available, the conclusion is that the availability of the SAPS predictions had no impact on the actual choice of ALMP. This conclusion holds with and without switchers, for the stock and for the flow sample, and in every region. For the flow sample, it even appears as if the treated complied less than the controls, but this difference is not statistically significant. The results in Appendix D.3 lead to similar conclusions for long-term compliance.

These results raise the question why caseworkers did not comply with the SAPS predictions. Two possible answers stand out: First, caseworkers make their own predictions and it could have happened that these largely provided the same

Table 2: Short-Term Compliance of Caseworkers in their Choice of ALMP

| | Observations (number of unemployed) | | Short term compliance | | | |
| | | | Highest prediction (Definition 1) | | Set of highest predictions (Definition 2) | |
	Treatment	Control	Treatment	Control	Treatment	Control
Including switchers	16566	14977	0.12	0.12	0.29	0.29
Without switchers	15701	14155	0.12	0.12	0.29	0.30
Stock sample	9844	8862	0.11	0.11	0.29	0.29
Flow sample	5857	5293	0.13	0.14	0.28	0.30
Basel	3528	3514	0.20	0.22	0.44	0.45
Bern	8484	7458	0.10	0.09	0.26	0.26
Geneva	1843	1437	0.04	0.05	0.14	0.14
St.Gallen	1578	1250	0.11	0.12	0.24	0.26
Zürich	2111	1933	0.08	0.07	0.19	0.18

Note: Unit of observation is the jobseeker. All rows (except the row labelled Geneva) are *without* Geneva. All rows (except for the first one) are *without* switchers. Definition 1 refers to the programme with the highest SAPS prediction. Definition 2 refers to all programmes, which belong to the MCB set of statistically significantly better programmes.

recommendations as the SAPS predictions. Second, caseworkers may have more or less ignored the econometric predictions either because of confidence in the superiority of their own judgements or because of unwillingness to comply with an external tool that could pose a potential threat to their future autonomy and discretion if introduced nationwide (perhaps with more binding predictions). The first of these answers can be ruled out as the SAPS predictions differed clearly from the choices made by the control caseworkers. The patterns recommended by SAPS and those realized by the caseworkers without accesses to the econometric predictions are very different. (See Table D.7 in Appendix D.3.)

4.3 The Impact of Downloading the SAPS Predictions on ALMP and Employment

Table 2 showed that on average we observe only very little or zero differences between treated and control groups in the realized choices regarding programme participation. One reason for this could be that caseworkers disregarded the SAPS predictions from the beginning by not even looking at them. Since the

SAPS predictions were made available via the Internet, we could monitor at what time exactly a caseworker inspected the predictions for a particular jobseeker. In a first instance, we examine how intensively caseworkers have made use of the statistical system to download predictions from the Internet, and whether this depended on characteristics of the caseworkers or the jobseeker. (Details can be found in Appendix D.2.) We observed that 21 of the caseworkers (= 15%) never examined the predictions at all, and that for many jobseekers SAPS predictions were never downloaded. We can thus partition the treatment group into jobseekers for whom predictions were never viewed and for those for whom they were examined at least once. It is hard to imagine that SAPS could have had an effect on those jobseekers for whom the predictions were never examined.

We could thus compare compliance and the choice of ALMP for those for whom SAPS predictions had been viewed versus those for whom they had not been inspected. Because the decision to download SAPS predictions is at the discretion of the caseworker and likely to be endogenous, such analyses could be affected by selection bias. We therefore perform instrumental variables regression of the effect of downloading SAPS predictions on compliance. We define for each jobseeker whether his SAPS predictions were ever downloaded during the field study. For a jobseeker of the control group, naturally, this never happens. It seems safe to assume that a caseworker who did *not* download the predictions is probably unlikely to comply more with the predictions than a caseworker of the control group. We can thus use the randomization as an instrument, denoted by Z, for the download of the predictions (treatment D) to estimate the impact of seeing the predictions on compliance status Y. This setup is thus very similar to the concept of Imbens and Angrist (1994). Since it was impossible for the members of the control group to download the predictions the monotonicity condition of Imbens and Angrist (1994) is satisfied by definition, and this also means that the local average treatment effect (LATE) is the same as the average treatment effect on the treated (ATET) because the treated are the compliers.[14] The exclusion restriction is our assumption that unemployed whose caseworker is in the treatment group but not downloading the predictions are not affected by their status as being randomized in, compared to unemployed with caseworker randomized to the control group.

Table 3 presents the impact of downloading the predictions on short-term compliance according to Definitions 1 and 2. We show IV estimates without control variables (i.e. Wald estimates) and with several control variables (2SLS), which

---

14   This is because there are no always-treated, in the language of Imbens and Angrist (1994).

Table 3: Impact of Downloading Predictions on Short-Term Compliance

Definition	All	Basel	Bern	St.Gallen	Zürich
Stock sample, with control variables					
1	0.00	−0.01	0.01	0.00	0.03
2	0.00	−0.02	0.03	−0.04	0.02
Flow sample, with control variables					
1	−0.02	−0.11	0.02	−0.05	0.00
2	−0.05	0.03	−0.02	−0.16	−0.04
Stock sample, without control variables					
1	0.00	−0.01	0.01	−0.01	0.02
2	0.00	−0.01	0.02	−0.07	0.02
Flow sample, without control variables					
1	−0.03	−0.15**	0.01	−0.03	0.01
2	−0.05	−0.01	−0.02	−0.11	−0.03

Note: Instrumental variable estimates of the effect of downloading the predictions on compliance, where compliance is measured in the short term, i.e. within 90 days. (Geneva is not included since predictions were hardly ever downloaded.) Standard errors are clustered by caseworkers. Significance at the 1%, 5% and 10% level is marked with ***, ** and *, respectively. The following characteristics of the jobseeker are used as control variables: female, age, foreigner with yearly permit, foreigner with residence permit, mother tongue neither German nor French, family size, insured earnings, qualification, employability rating, looking for a part-time job, duration of unemployment, unemployment spells in last two years.

may produce more precise estimates due to efficiency gains. The first stage regression (not shown) is highly significant in both cases (indicating that the instrument is not weak). Table 3 shows, however, that all estimates of the effects are *insignificant*. The only exception is a *negative* effect in the flow sample in Basel. The effects in the stock sample are zero throughout.

In Appendix D.4 the corresponding estimates for the long-term compliance are given. The results are similar for Definitions 1 and 2, with the negative effect in Basel now only being marginally significant at the 10% level in one of the two variants considered. Some evidence for a negative effect in St.Gallen appears as well. The two additional definitions of long-term compliance (Definition 3 and 4) examined in Appendix D.4 show some indications of a positive effect in Zürich, which is not very stable, though. Given that we observe no significant effect for Definitions 1 and 2, this result has to be interpreted with care as

Definition 3 may be afflicted by some selection bias (see Appendix D.3 for details on defining compliance). For the other regions, all effects are insignificant.

Hence, these results corroborate the finding that caseworkers did not adjust their selection of ALMP to the econometric predictions provided by SAPS. In additional analyses (not shown here) we also examined IV estimates of downloading on the employment probabilities of the jobseekers. Not surprisingly, no clear and significant pattern was found.

5. Conclusions

A randomized experiment was conducted in Switzerland to evaluate the potential of a statistical targeting system to assist caseworkers in choosing active labour market programmes for their unemployed clients. The potential employment outcomes were predicted for each unemployed person based on a large administrative dataset. The experiment was designed such that caseworkers retained full discretion about the choice of labour market programmes for their unemployed clients. The evaluation results showed that caseworkers largely ignored the statistical support system. No significant differences in their choices of labour market programmes could be discerned vis-à-vis the experimental control group. Caseworkers either decided to ignore the system out of various reasons, or they were overly confident in their own experience and considered econometric estimates as inferior.

Profiling and targeting of ALMP is a hotly debated topic and several countries including the UK, Germany, Denmark, Finland, and Sweden are currently piloting such systems or considering their introduction. Our evaluation results indicate that caseworker discretion may conflict with the provision of statistical targeting. In order to evaluate the impact of these systems, specific incentives may be required for caseworkers to comply. Bonus payments for caseworkers who complied with the experiment may be one example. These positive incentives should compensate for potential negative effects due to being unfamiliar with the system or being uncertain about its merits during the pilot phase. Although such incentives might often not be appropriate in a fully nationwide implementation of the system, such monetary payments during the pilot study could help a lot to permit an evaluation of the potential of targeting systems.

Appendix

A.   Data

For the pilot study, detailed data on previous clients was required for the estimation of the coefficients of the SAPS-system. Furthermore, data was needed for the clients during the field study and their follow-up information on assignments to ALMP and subsequent employment outcomes.

The estimation of the SAPS-system was based on the entire population of individuals (aged 25 to 55) who registered as jobseekers at an employment office anytime during January 2001 to December 2003. For these 460442 persons, detailed information from the unemployment insurance information system (AVAM/ASAL) was available from January 1998 to December 2004. This data set was matched with the complete monthly information from the social security and pensions system (AHV) for the period January 1990 to December 2002. These combined data sources contain very detailed information on registration and de-registration of unemployment, benefit payments and sanctions, participation in ALMP, eleven years employment histories with monthly information on earnings and employment status (employed, unemployed, non-employed, self-employed). Furthermore, they contain information on socioeconomic characteristics including qualification, education, language skills (mother tongue, proficiency of foreign languages), job position, experience, profession, industry, and an employability rating by the caseworker, among other variables.

During the field study in 2005, the information from the unemployment insurance information system (AVAM/ASAL) was delivered to us biweekly for all currently registered jobseekers.

Finally, in February/March 2007 we received the data from the unemployment insurance information system until the end of December 2006. This permits us to follow up each participant of the field study for at least 12 months.

B.   Further Details on the SAPS Predictions

The SAPS predictions are based on estimates of the expected potential outcomes

$$E[Y^0_{i,t+\tau} \mid X_{it}],\dots,E[Y^R_{i,t+\tau} \mid X_{it}]$$

which are used to estimate the expected optimal treatment:

$$r^*(X_{it}) = \underset{r\in\{0,\dots,R\}}{\arg\max}\; E[Y^r \mid X = X_{it}].$$

The SAPS system provided estimates of the employment chances

$$E[Y^0_{i,t+\tau}\mid X_{it}],\dots,\; E[Y^R_{i,t+\tau}\mid X_{it}]$$

and of the optimal treatment r^*, based on a rich set of covariates $X_{i,t}$, which includes characteristics of the unemployed person, such as age, gender, family composition, education, language skills, qualifications, job experience, past employment and earnings histories, previous participation in programmes etc., and of the local labour market. Several of these covariates are time varying, e.g. the current duration of the unemployment spell or the number of vacancies in the local labour market.

For the estimation of the system, an even larger set of covariates was taken into account to eliminate potential selection bias, see Frölich (2008). Since selection bias may be more of a concern for the young and the older individuals, the SAPS system is restricted in this pilot version to the 25 to 55 year old jobseekers. (For the younger jobseekers, detailed and long employment histories are often not yet available. For the older jobseekers good health data would be helpful e.g. to assess their labour market attachment or early retirement options.)

In addition to these estimates of expected employment chances, the SAPS system also provides estimates of statistical precision, which are conveyed to the caseworker in a simple and accessible way. Using Multiple Comparison with the Best procedures (MCB), the available programmes are separated into three categories: 'good', 'intermediate', and 'bad' treatments. The information provided to the caseworker to assist the treatment choice for jobseeker i is in the following form and was made accessible via an Internet application developed for the field study (see Table B.1).

Programmes that are statistically significantly better than the others are marked as **<u>bold underlined</u>**, whereas intermediate programmes are marked in

Table B.1: Example of Individual SAPS Predictions

Expected number of months in stable employment in the following 12 months for individual i if initiating a labour market programme now or soon:

No programme	**6.7**
Job search and personality courses	2.7
Language skills training	4.1
Computer skills training	**6.1**
Further training	**5.7**
Employment programmes	3.0

bold and the worst programmes are not marked. (In the pilot study, the best programmes were marked on the screen in green and the worst in red.) Caseworkers were advised to choose a programme out of the *set of statistically best programmes*, with a slight preference to be given to the programme with the highest estimated employment chances. In the example above, the set of best programmes contained the options "no programme" and "computer skills training".

Generally, the cardinality of this set depends on the covariates $X_{i,t}$. For some jobseekers, there was only one statistically significantly best programme, whereas for others the set of best programmes might contain three or four, or even all programmes.[15] This was intended to show the caseworker that the information that the statistical system can provide varies across jobseekers and that the caseworker should trust the SAPS predictions more if they were very precise and consider other considerations when they were very imprecise.

C. *Labour Market Programmes in Switzerland*

Many different programmes are available in Switzerland (and these programmes might vary somewhat from region to region). The official classification distinguishes 43 different types, of which most are training or employment programmes. For various reasons explained below these programmes were grouped into a few broader categories. The exact definition of the groups varied slightly

---

15 The confidence level for the statistical inference was also randomized across caseworkers but this did not seem to affect their behaviour at all.

from region to region and the following discussion focuses on Basel city. For the region Basel the ALMP are categorized into six ($R + 1 = 6$) different groups:

Table C.1: Labour Market Programmes in Basel

No programme
Job search and personality courses
Language skills training
Computer skills training
Further training
Employment programmes

The first category 'No programme' means that the jobseeker is not allocated to any ALMP in this month, but leaving the option for the future, if still unemployed then. This category could therefore also be labelled as 'waiting' or 'no programme now but perhaps later'. This has to be distinguished from a treatment 'no programme at all' or 'no programme for the next 12 months' or 'no programme for the entire unemployment spell'. Such a programme does not exist in the above list out of two reasons: First, forgoing the option to choose a labour market programme later is not really a sensible choice for a caseworker. The caseworker meets the jobseeker about once a month and decides about actions to be taken then. Sequential plans may be developed but at every meeting, the latest information and events are incorporated to update such plans. Second, identifying the effect of a treatment 'no programme for the next 12 months' is more difficult than for a treatment 'no programme now but perhaps later' because of the dynamic nature of the job search. When examining previous participants in 'no programme for the next 12 months', many of them had been lucky enough to find a job before a programme had been assigned. Hence, this group may contain a larger proportion of good risks or individuals successful in the job search. For further discussion, see Fredriksson and Johansson (2003).

The categories two to six contain active programmes.[16] The second category consists of a variety of often short, basic courses, including training in

---

16 Only courses of at least five days duration are included. Shorter courses are included in the no programme category. Such may be short evening courses that provide information on the duties and rights of unemployed or language proficiency tests for assessing the need for a language course or its appropriate level.

effective job search strategies and resume writing and more intensive personality courses, which provide psychological backing for handling the shock of becoming unemployed and coaching in developing new perspectives to entering the labour market. These courses may be tailored to different groups (manual workers, management) and offered in different languages.

The third category contains language and communication skills training for foreigners (including alphabetization courses, basic skills in dealing with Swiss administrations and vocational language courses for low educated foreigners)[17] as well as courses in foreign languages at different levels. Category 4, computer training, refers mostly to general courses in office applications such as word processing and spreadsheet calculations, but also stock keeping and order management software. The fifth category consists of further training in the jobseeker's occupation. Its duration is usually between one week and two months. (Retraining to a new profession is not offered by Swiss ALMP.)

The sixth category consists of subsidized employment programmes or job creation schemes in a sheltered labour market, usually of three to six months duration. This includes activities in cantonal and municipal administrations (including hospitals, kindergartens, schools, and nursing homes) and non-regular workplaces in charitable, cultural, recycling, environmental protection or other non-profit organizations. Internships are also included in this category.

Given the large number of active labour market programmes available in Switzerland the above grouping into only 5 broad categories may appear rather rough. There are several reasons for not choosing very narrow categories, though. One reason is statistical precision in that the number of observations available in the dataset would be very small for some courses. However, there are also more substantial issues. First, all of the R available categories should make sense for *every* jobseeker. If one of the categories was defined as a language course for foreigners, it would not be a reasonable option for a Swiss jobseeker and no predictions should be made because such a programme would be dismissed from the outset. The choice set $\{0, …, 5\}$ would thus depend on the characteristics X_{it} and would have to be treated as a function of X_{it}, which would complicate the implementation. By defining a category *language skills training* which includes German, French and foreign language courses, this category becomes feasible for every jobseeker, and the X_{it} characteristics (e.g. mother tongue, profession) define which type of language course or further training is appropriate.

---

17 Learning occupation specific vocabulary e.g. in the construction or hotel and restaurant industry.

A second reason is that the caseworker may actually have much better information for choosing the exact course out of a broader category. The statistical system may be able to estimate how much the labour market values different types of training, but cannot recommend whether an advanced or intermediate English course would be more appropriate. The caseworker may also know better about local waiting lists or supply constraints that are to be taken into account when allocating a course.

Third, in the pilot study employment predictions are made for the year 2005/06 based on data on participants of the years 2001 to 2003. During these years, some of these courses have been modified and providers have changed in several details. However, the broader structure of these programmes remained largely unchanged. Therefore, we do not want to define categories too narrowly, as specific courses may be rather different today.[18]

D. Evaluation of the Field Study

D.1 Sample

In this appendix additional material about the evaluation of the experiment is given. The following table shows the number of treated and control jobseekers in the five regions.

A jobseeker is defined as treated or control according to the treatment status of his *first* caseworker during the field study. For some jobseekers, a change in their caseworker happened, where the new caseworker might have a different treatment status. The occurrence of such a switching is examined in the following table. Examining the stock sample, one can see that for about 800 of the treated and control jobseekers their treatment status changes over time. Of those jobseekers whose caseworkers were in the treatment group at the beginning of the field study, 238 moved to a caseworker of the control group and 556 moved to a caseworker who belongs neither to the SAPS treatment nor to the control group.[19] These jobseekers

---

18   The above categories contain only programmes that a caseworker can actively assign. The Swiss labour market policy also provides a few other instruments, such as subsidies for temporary jobs (interim jobs), regular jobs (settling-in allowances), and self-employment assistance. These are not included in the statistical system since the former are largely contingent upon that a job has already been found (and thus cannot be assigned directly by the caseworker) and since the occurrence of self-employment assistance is relatively rare and the selection problem more difficult to handle.

19   These caseworkers did not participate in the randomization or started working for the employment office after the randomization.

Table D.1: Sample Sizes

	Stock sample 22758		Flow sample 12632	
	treatment	control	treatment	control
All	12079	10679	6634	5998
All without Geneva	10401	9472	6165	5505
Basel	2404	2368	1158	1202
Bern	5501	4805	3474	3004
Geneva	1678	1207	469	493
St.Gallen	1078	960	757	549
Zürich	1418	1339	776	750

Note: Number of treated and control jobseekers in the five regions. The first row refers to the entire sample. The second row refers to all regions except Geneva.

started as being treated but then received no predictions anymore. (A treatment effect of SAPS might thus be diminished for this group.) Note that there are only 16 cases changing treatment status more than once over time. For those jobseekers who started being in the control group, 448 moved to a caseworker who belonged neither to the treatment nor to the control group. For these jobseekers, predictions never became available. The only group that could raise concern are those 345 observations who started as controls and ended up in the treatment group. These could bias a treatment effect downward. Nevertheless, they represent only about 3% of the controls.[20] For the flow sample, these figures are lower since a caseworker change usually takes place only after several months of unsuccessful job search attempts. Overall, the numbers of switchers are not very large to expect large impacts on estimated treatment effects. In fact, in the main analyses most estimates turn out to be very similar with and without switchers.

In some additional analyses (not shown here), we also examined the switchers in more detail. The occurrence of caseworker changes was more frequent in Geneva and in St.Gallen (where there was substantial staff turnover during the field study). Caseworker changes happened less frequently in Basel, Bern, and

---

20   There is also a group of jobseekers who started as being neither treated nor control and entered later into the treatment or control group. These, however, are eliminated from the sample.

Table D.2: Change in Treatment Status over Time

	Stock sample		Flow sample	
	Treatment status according to the time-invariant definition			
	treatment	control	treatment	control
	22 758		12 632	
All	12 079	10 679	6 634	5 998
Time constant	11 269	9 875	6 276	5 707
Switchers	810 (6.7%)	804 (7.5%)	358 (5.4%)	291 (4.9%)
into treatment		345		143
into control	238		149	
to other	556	448	205	145
several changes	16	11	4	3

Note: Stock and flow sample defined according to the time-invariant definition, where treatment status of the jobseeker is defined by the treatment status of his first caseworker.

Zürich. An analysis of the individual characteristics of the jobseekers showed that the switchers had poorer labour market chances than the non-switchers. Nevertheless, those who switched into treatment had similar characteristics as those who switched into control. Similarly, those who switched out of treatment also had similar characteristics to those who switched out of control. Hence, there does not seem to be a systematic difference between the switchers. One could have been concerned that employment office managers might have allocated jobseekers, for whom a caseworker change was due, selectively into the treatment or control group. However, this is not supported by the data. Hence, leaving out the switchers from the analysis should lead to less contaminated impact estimates without introducing selection bias.

Download of SAPS Predictions

The 142 caseworkers in the treatment group could download the SAPS predictions anytime for any of their clients. Twenty-one caseworkers never used this option at all. The following discussion provides some information on the frequency of downloading the SAPS predictions.

Overall, for about 37% of all jobseekers in the treatment group their SAPS predictions were examined *at least once*. Hence, for nearly two thirds they were

never inspected. For the 142 caseworkers the following table shows for how many of their clients the SAPS predictions were viewed at least once. The average (median) caseworker counselled during the entire field study a total number of 163 jobseekers. The total number of cases per caseworker varied from a minimum of 8 to a maximum of 283 clients. This variation is partly due to different caseloads but also due to different lengths of the field study, which did not start in all offices at the same time and ended earlier for those caseworkers who retired. It is also due to different labour market conditions, e.g. the pace of job turnover. The average caseworker downloaded SAPS predictions for 51 different clients, whereas the most active caseworker examined predictions for 174 clients. The average caseworker examined the SAPS predictions for 33% of his clients, with the most active caseworker even viewing the predictions for 83% of his clients. The frequency of downloading the predictions was largest in St.Gallen and Zürich. In Geneva, on the other hand, the SAPS predictions were downloaded on average for less than 1% of all clients (not shown). In other words, SAPS was almost completely ignored in Geneva.

Table D.3: Frequency of Download of SAPS Predictions per Caseworker (All Regions without Geneva)

	25% quantile	Median	75% quantile	Minimum	Maximum
Number of jobseekers for whom predictions were downloaded	8	51	78	0	174
Total number of jobseekers during field study	129	163	188	8	283
Fraction of jobseekers for whom predictions were downloaded	0.08	0.33	0.57	0	0.83
in Basel	0.09	0.40	0.56	0	0.81
in Bern	0.13	0.37	0.56	0	0.83
in Geneva	0.00	0.00	0.01	0	0.25
in St.Gallen	0.17	0.54	0.62	0.02	0.72
in Zürich	0.18	0.51	0.67	0	0.79

Note: 142 caseworkers. All rows (except the row labelled Geneva) are *without* Geneva. For each variable in column 1 the 0.25, median and 0.75 quantile and the minimum and maximum over the 142 caseworkers (without those in Geneva) is given.

Table D.4: Probability of Download as a Function of Jobseeker Characteristics

Dependent variable	With switchers	Without switchers					
Download of prediction	All without Geneva	All without Geneva	Basel	Bern	Geneva	St.Gallen	Zürich
Female jobseeker	−0.036	−0.036	0.031	−0.15***	0.158	0.059	0.17*
Age/100	0.257	0.288	1.542	−0.96	6.087*	3.68**	0.66
Age/100 squared	0.31	0.329	−1.41	1.9	−7.501*	−4.325**	−0.03
Foreigner with yearly permit (B)	−0.055	−0.045	0.06	−0.14**	−0.331***	−0.007	0.08
Foreigner with residence permit (C)	−0.008	−0.015	0.013	−0.05	−0.241***	0.028	−0.05
Qualification (reference: apprenticeship/degree)							
No qualification	−0.035	−0.036	0.043	−0.1	−0.538**	−0.015	0.15
Short apprenticeship	0.126**	0.127**	−0.08	0.09	−0.304*	0.007	0.26*
Apprenticeship or degree that is not officially recognized in Switzerland	−0.002	−0.01	−0.09	0.01	−0.071	0.126	0.07
Employability medium	0.129	0.148	0.067	0.19	−0.467**	0.668***	0.08
Employability difficult	0.1	0.11	0.215	0.18	−0.256*	0.381*	0.08
Looking for part-time job	−0.122***	−0.133***	−0.1	−0.14**	−0.352	−0.096	0.05
Current duration of unemployment	0.00	0.00	0.00	0.00**	0.00	0.001***	−0.00**
Number of unemployment spells in last 2 years	0.209***	0.224***	0.328***	0.16***	0.189	0.531**	0.25*
In flow sample	−0.48***	−0.475***	−0.62***	−0.45***	−0.514***	−0.4***	−0.42***
Constant	−0.325*	−0.348*	−0.64*	−0.18	−2.275**	−1.382***	−0.22
N	16 563	15 699	3 528	8 482	1 843	1 578	2 111

Note: Probit regression. All jobseekers of the treatment group (without Geneva). Dependent variable: whether, for the particular jobseeker, SAPS predictions have been downloaded at least once. Standard errors are clustered by caseworkers. Significance at the 1%, 5% and 10% level is marked with ***, **, and *, respectively.

In the Tables D.4 and D.5, we examine whether the probability of downloading the predictions depends on the jobseeker characteristics and/or on the caseworker characteristics. Table D.4 shows the probit regression of download on several jobseeker characteristics. The estimation results for Geneva should be interpreted with caution since the SAPS predictions were hardly ever downloaded. The main pattern observed is a higher likelihood of examining SAPS predictions for those who had been more unemployed in the last two years, and a lower likelihood for those jobseekers in the flow sample. The latter may be due to a somewhat declining interest on the side of the caseworkers during the pilot study.

Table D.5 includes also caseworker characteristics additionally in the regression. The caseworker characteristics are mostly insignificant and their coefficients do not display a clear pattern across regions. For the jobseeker characteristics it again appears that the number of unemployment spells and the current duration of unemployment has a positive effect on the download probability. The negative coefficient for the flow sample remains. The variable 'short apprenticeship' (=Anlehre) is positively significant in both tables. In Switzerland, a 'short apprenticeship' is of two years duration with reduced skill requirements, and is distinct from a 'regular apprenticeship' which is of three years duration and more demanding. Jobseekers with a 'short apprenticeship' are thus not unqualified but signal a clear qualification gap, which could perhaps partly be filled with ALMP. Overall, the occurrence of a download of the SAPS prediction depended more on the jobseeker characteristics than those of the caseworker, but the overall patterns are not very strong.

In additional analysis (not shown here), we also examined to which extent the dynamic nature of the predictions was used: the predictions were updated every two weeks for every jobseeker still registered. Hence, caseworkers should inspect the predictions repeatedly for the same client, ideally before every counselling meeting. Overall, it is observed that the frequency of downloading slowly declined during the field study and that not much use was made of the regular updating of the SAPS predictions.

D.2 Long-Term Compliance with the SAPS Predictions

In Table 2, we observed that the incidence of "compliance" with the SAPS predictions was almost identical in the treatment and the control group. Compliance was defined if SAPS predictions and actual treatment choice coincided within the subsequent 90 days. In the following, we consider alternative definitions of compliance, which nevertheless all lead to the same conclusion that differences in compliance between treated and control are (close to) zero.

Table D.5: Probability of Download as a Function of Jobseeker and Caseworker Characteristics

Dependent variable	With switchers	Without switchers				
Download of prediction	All without Geneva	All without Geneva	Basel	Bern	St.Gallen	Zürich
Jobseeker characteristics						
Female jobseeker	−0.021	−0.019	0.043	−0.089*	0.1*	0.101
Age/100	0.408	0.444	1.262	−0.908	4.362**	1.803
Age/100 squared	0.131	0.145	−1.146	1.857	−5.13*	−1.443
Foreigner with yearly permit (B)	−0.05	−0.041	0.075	−0.123**	−0.028	0.035
Foreigner with residence permit (C)	−0.009	−0.017	0.013	−0.061	0.064	−0.046
Qualification (reference: apprenticeship/degree)						
No qualification	−0.038	−0.044	0.074*	−0.125*	0.017	0.148**
Short apprenticeship	0.128**	0.129**	−0.027	0.084	−0.181	0.233*
Apprenticeship or degree that is not officially recognized in Switzerland	−0.001	−0.009	−0.101	0.029	0.071	0.009
Employability medium	0.127	0.146	0.037	0.177	0.407***	−0.033
Employability difficult	0.071	0.081	0.126	0.123	0.391	0.144
Looking for part-time job	−0.118***	−0.128***	−0.089	−0.097**	−0.288*	0.006
Current duration of unemployment in days	0.00	0.00	0.00	0.001**	0.001**	−0.001***
Number of unemployment spells in last 2 years	0.215***	0.229***	0.303***	0.174***	0.561**	0.271**
In flow sample	−0.48***	−0.474***	−0.65***	−0.46***	−0.421***	−0.485***

Table D.5 continued

Dependent variable	With switchers	Without switchers				
Download of prediction	All without Geneva	All without Geneva	Basel	Bern	St.Gallen	Zürich
Caseworker characteristics						
Female caseworker	−0.206	−0.201	−0.317	−0.415**	0.644*	0.582*
Age: 25 to 40 years	−0.135	−0.079	−0.303	0.117	−0.61***	0.09
41 to 46 years	0.118	0.12	−0.17	0.17	−1.579***	0.305
46 to 56 years	0.271	0.325*	0.367	0.11	−0.262	0.931**
age missing	0.114	0.099	−0.294	0.043	0.151	n/a
Constant	−0.357*	−0.4*	−0.334	−0.101	−1.102***	−0.983*
N	16 563	15 698	3 528	8 481	1 578	2 110

Note: Probit regression. All jobseekers of the treatment group (without Geneva). Dependent variable: whether, for the particular jobseeker, SAPS predictions have been downloaded at least once. Standard errors are clustered by caseworkers. Significance at the 1%, 5% and 10% level is marked with ***, ** and *, respectively.

In the following definitions of compliance, again the first SAPS prediction is examined and compared to the actual assignment to ALMP, where the entire period after the SAPS prediction *until December 2006* is considered. In addition, the first *three* assigned ALMP are examined.[21] Hence, a caseworker who assigned first a non-recommended ALMP but afterwards as a second (or third) ALMP also a recommended programme would be considered as complying, even if the second or third programme was assigned only many months later. Hence, we permit a delayed compliance in that a caseworker might have initiated a different programme first with the option to consider the SAPS suggestions later. This thus captures the *long-term* effect of the SAPS predictions on the choice of ALMP.

The following table shows the results for 8 different definitions. To structure these different definitions we distinguish between 4 definitions and 2 versions. The 2 versions differ in how they define when a jobseeker has realized the option "no programme".

In Version A a jobseeker is considered to have *received* "no programme" if he was not assigned to any ALMP at all until the end of our observation period in December 2006. (If he was assigned to at least one ALMP, he is considered to have not received "no programme".) In Version B a jobseeker is considered to have *received* "no programme" if he was assigned to *at most two* ALMP until the end of our observation period in December 2006. I.e. he is considered to have received "no programme" *in addition to* the other ALMP that he participated in. If he was assigned to three or more ALMP, he is considered to have not received "no programme". In other words, "no programme" in Version A means no ALMP at all, whereas in Version B it means at most two ALMP, but not more.

The four different definitions differ in how they define whether SAPS predictions and actual allocations coincided. In Definition 1 compliance is defined if the programme with the highest SAPS predictions coincided with any of the first three assigned programmes (including the "no programme" option). In Definition 2 compliance is defined if *any* of the programmes that are in the MCB set of best programmes coincided with *any* of the first three assigned programmes

---

21   Only ALMP *after* the first SAPS-prediction are considered. For the stock sample, this corresponds with the start of the field study. For a jobseeker of the flow sample, this corresponds to the first date of uploading predictions for this individual. We focus on the *first* SAPS prediction being made, because compliance with later predictions are harder to measure as there might be a delayed effect from previous predictions. In addition, the caseworkers' interest and attention seems to have been highest in the early phase of the field study.

(including the "no programme" option).[22] Although this definition of compliance is extremely wide, it was satisfied only for 29% (Version A) or 35% (Version B) of all jobseekers, as Table D.6 shows. Examining the results for treated and control we observe no systematic difference in compliance rates. If any difference is discernible at all, it even points towards slightly less compliance of the treated caseworkers. As before, the only exception is Geneva. Since SAPS predictions were hardly ever downloaded in Geneva, this difference can only be the result of randomization not having been successful in Geneva in that it did not lead to a balanced control group. (The compliance rates for Basel are much higher than for the other regions, which is at least partly due to the smaller number of ALMP categories in Basel, which were only 6 for Basel and 7 for the other regions, including the no participation option.)

Before concluding that no compliance difference exists, we examined two additional definitions of compliance. One could imagine that our findings might have to do with the fact that we treated the "no programme" option just as another treatment option in the SAPS predictions. However, it could well be that caseworkers consider this option as something very different from an active choice of an ALMP.

First, one could imagine that caseworkers first decide (i.e. without consulting SAPS) whether using ALMP or not is most appropriate for a particular client. Only if they conclude that an ALMP should be assigned, they might wish to examine the SAPS predictions. In this case, we would like to measure compliance only in the second step of this two-stage decision process. We therefore consider as compliance if the jobseeker was assigned to the programme with the *largest SAPS prediction, given that an ALMP was assigned* (Definition 3). In other words, for Definition 3 all jobseekers who never received any ALMP are deleted, and for the remaining jobseekers, we apply Definition 1: Compliance is defined if the programme with the highest SAPS prediction coincided with any of the first three assigned programmes (not including the "no programme" option).

The results are shown in Table D.6. (Because the definition is the same for Version A and B, no numbers for Version B Definition 3 are given.) The note below Table D.6 gives the number of observations who are not deleted for Definition 3,

---

22 Note that, for Version B, these definitions include the case of short-term compliance as discussed in Table 2. Version A of long-term compliance, however, is neither strictly weaker nor stronger than short-term compliance. It is weaker in the sense that long-term compliance considers the first three programmes assigned. It is stricter in the sense that "no programme" requires no ALMP until the end of 2006, whereas short-term compliance only requires no ALMP for the subsequent 90 days.

which are roughly a third of the total sample. If the decision to assign any ALMP at all or not was already influenced by SAPS or related to some unobserved characteristics, this could be a selected sample and might incorporate some selection bias. Again, there seems to be no difference between treatment and control group, perhaps with the exception of Zürich, where compliance according to this definition was higher for the treated than for the compliers.

For the second alternative definition (Definition 4), one could imagine that caseworkers may have considered the SAPS predictions as too much tilted towards active ALMP. Therefore, caseworkers may have followed the SAPS predictions in general but preferred to choose the "no programme" option more often. (The target variable of the SAPS project ignored programme costs entirely and due to its focus on stable employment, penalized short job spells; both aspects could have been at odds with the preferences of (some of) the caseworkers.)

In Definition 4, we do not eliminate those cases where no ALMP has been chosen but consider instead the "no programme" option to be always among the set of best options. Hence, we artificially augment the MCB set of best programme with the "no programme" option. Compliance is then defined, analogously to Definition 2, if the jobseeker was assigned to a programme out of the *set of significantly best programmes including "no programme"* (Definition 4).[23]

As Table D.6 shows, there is again no systematic difference in compliance rates between treated and control. Only in Zürich a small difference can be observed, which is not significant. (The overall much higher compliance rate of about 80% is due to the artificial augmentation with the no programme option, since many jobseekers did not receive any ALMP at all.)

Hence, whichever definition we look at, the conclusion is always nearly the same. Only in Zürich a (insignificant) difference can be observed for Definitions 3 and 4, but none for the, more plausible, Definitions 1 and 2. For the other regions, no differences are discerned and sometimes it appears even as if there is less compliance among treated than controls. Table D.6 thus suggests that the availability of the SAPS-predictions seems to have been ignored almost completely by the caseworkers, at least in the sense that their choices of ALMP are not significantly different from those of the control group.

---

23 More precisely, compliance according to Version A is defined if either never an ALMP has been assigned or if any of the first three assigned ALMP coincided with the MCB set of best programmes. Version B is complied with if either Version A is satisfied or if less than three ALMP have been assigned until December 2006 *and* "no programme" belonged to the MCB set of best programmes according to the SAPS predictions.

Table D.6: Long-Term Compliance of Caseworkers in Their Choice of ALMP

	Highest prediction (Definition 1)		Set of highest predictions (Definition 2)		Highest prediction, only if an ALMP was assigned (Definition 3)		Set of highest predictions including no ALMP (Definition 4)	
	Treat-ment	Control	Treat-ment	Control	Treat-ment	Control	Treat-ment	Control
Long-term compliance (Version A)								
Including switchers	0.12	0.13	0.29	0.29	0.27	0.26	0.75	0.74
Without switchers	0.13	0.13	0.29	0.29	0.27	0.26	0.75	0.74
Stock sample	0.12	0.12	0.30	0.30	0.29	0.29	0.78	0.77
Flow sample	0.13	0.14	0.28	0.28	0.23	0.23	0.71	0.71
Basel	0.19	0.19	0.41	0.41	0.28	0.30	0.77	0.77
Bern	0.10	0.10	0.27	0.26	0.25	0.25	0.74	0.73
Geneva	0.11	0.15	0.23	0.26	0.39	0.44	0.78	0.74
St.Gallen	0.12	0.13	0.25	0.27	0.26	0.26	0.72	0.70
Zürich	0.10	0.10	0.22	0.21	0.34	0.27	0.80	0.78
Long-term compliance (Version B)								
Including switchers	0.16	0.17	0.35	0.36			0.81	0.80
Without switchers	0.16	0.17	0.35	0.36			0.81	0.81
Stock sample	0.15	0.15	0.35	0.36			0.84	0.83
Flow sample	0.18	0.19	0.35	0.36	same as above		0.78	0.78
Basel	0.25	0.27	0.51	0.52			0.87	0.87
Bern	0.14	0.13	0.32	0.32			0.80	0.79
Geneva	0.11	0.16	0.25	0.28			0.80	0.76
St.Gallen	0.16	0.18	0.30	0.34			0.76	0.77
Zürich	0.12	0.11	0.25	0.23			0.83	0.80

Note: Unit of observation is the jobseeker. All rows (except the row labelled Geneva) are *without* Geneva. All rows (except for the first one) are *without* switchers. The number of observations for Definition 3 is 5689 in the treatment and 5333 in the control group. Without switchers, these are 5293 and 4929. In stock sample: 3075 and 2978. In flow sample: 2218 and 2021. In Basel, Bern, Geneva, St.Gallen and Zürich these numbers are 1110 and 1176, 2960 and 2653, 679 and 568, 599 and 510, 624 and 590, respectively.

Table D.7 compares the patterns of ALMP that were recommended by SAPS with those actually assigned in the treatment and control group. It shows, first, that there were no systematic differences between treated and control, and, second, that the SAPS recommendations were clearly different from those actually realized. Hence, the potential argument that the reason for why we did not observe any effect on the pilot study being the coincidence of the caseworkers' own predictions with those of SAPS is not supported.

Table D.7: Recommended and Realized Treatment Choices

| | First SAPS prediction | | Allocated programme after start of pilot study | | | |
| | Highest prediction | | First programme | | First three programmes | |
	Treatment	Control	Treatment	Control	Treatment	Control
Basel						
Number of observations	3528	3514	3528	3514	3528	3514
No ALMP	21.15	22.65	68.54	66.53	97.93	97.84
Basic courses (job search training, personality course)	11.45	12.07	10.12	12.81	10.23	12.86
Language course	9.47	9.36	3.26	3.22	3.74	3.84
Computer course	15.31	14.88	2.98	2.25	3.49	3.02
Further training	27.52	26.07	2.27	1.79	2.98	2.45
Employment programme	15.11	14.97	6.15	6.63	8.62	9.68
Other courses			6.69	6.77	8.76	8.48
Bern						
Number of observations	8484	7458	8484	7458	8484	7458
No ALMP	10.25	9.76	65.10	64.43	95.95	95.64
Basic courses (job search training, personality course)	1.98	1.78	10.58	10.53	10.89	10.78
Language course	12.28	12.24	5.32	5.36	6.22	6.30
Computer course	19.10	20.10	2.25	2.16	3.22	3.24
Further training	21.60	21.47	2.85	2.70	4.14	3.96
Employment programme (individual) / internship	28.58	27.25	2.31	2.44	3.14	3.35

Table D.7 continued

	First SAPS prediction		Allocated programme after start of pilot study			
	Highest prediction		First programme		First three programmes	
	Treatment	Control	Treatment	Control	Treatment	Control
Employment programme (collective) / training firm	6.20	7.40	4.96	4.83	7.44	6.74
Other courses			6.62	7.56	8.99	9.90
St.Gallen						
Number of observations	1578	1250	1578	1250	1578	1250
No ALMP	11.91	12.24	62.04	59.20	94.74	94.24
Basic courses (job search training)	6.34	5.92	11.09	9.28	11.22	9.44
Personality course	16.98	14.56	6.65	9.12	9.38	11.68
Language course	24.40	25.36	3.80	5.28	6.08	6.80
Computer course	7.22	8.16	2.66	2.96	4.82	4.88
Further training	30.42	30.96	0.32	0.40	0.51	1.12
Employment programme	2.72	2.80	8.87	9.36	13.94	15.28
Other courses			4.56	4.40	6.65	6.56
Zürich						
Number of observations	2111	1933	2111	1933	2111	1933
No ALMP	7.58	6.67	70.44	69.48	98.20	98.86
Basic courses (job search training)	7.67	6.41	15.11	15.26	15.11	15.26
Personality course	8.15	6.83	1.71	1.66	2.42	2.43
Language course	31.83	32.44	4.55	4.50	5.73	5.69
Computer course	12.08	15.73	1.71	1.60	2.65	2.38
Further training	16.96	16.14	0.90	0.57	1.18	0.72
Employment programme	15.73	15.78	4.50	4.86	6.73	7.09
Other courses			1.09	2.07	1.56	2.53

Note: All jobseekers, without switchers, without Geneva.

D.3 IV Estimates of Long-Term Compliance with the SAPS Predictions

Tables D.8 and D.9 provide the instrumental variable estimates for long-term compliance, and thus complements Table 3 which showed the corresponding results for short-term compliance.

Table D.8: Impact of Downloading Predictions on Long-Term Compliance (Version A)

Definition	All	Basel	Bern	St.Gallen	Zürich
Stock sample, with control variables					
1	0.00	0.02	0.00	0.02	0.02
2	0.00	0.00	0.02	−0.02	0.02
3	0.02	−0.06	0.03	0.02	0.10**
4	0.03*	0.01	0.04	0.03	0.05
Flow sample, with control variables					
1	−0.02	−0.01	0.01	−0.09	0.01
2	0.00	0.11	0.01	−0.09	0.01
3	0.00	−0.03	−0.04	−0.03	0.18
4	0.02	0.08	−0.03	0.06	0.11**
Stock sample, without control variables					
1	0.00	0.01	0.00	0.01	0.01
2	0.00	0.00	0.02	−0.04	0.02
3	0.02	−0.06	0.04	0.03	0.10**
4	0.04*	0.01	0.05	0.03	0.06*
Flow sample, without control variables					
1	−0.03	−0.02	−0.02	−0.08	0.01
2	0.00	0.10	0.00	−0.05	0.02
3	0.01	0.00	−0.03	0.00	0.18
4	0.01	0.08	−0.03	0.08	0.07

Note: Instrumental variable estimates of the effect of downloading the predictions on compliance, where compliance is measured in the long term (Version A). (Geneva is not included since predictions were hardly ever downloaded.) Standard errors are clustered by caseworkers. Significance at the 1%, 5% and 10% level is marked with ***, ** and *, respectively.

Table D.9: Impact of Downloading Predictions on Long-Term Compliance (Version B)

Definition	All	Basel	Bern	St.Gallen	Zürich
Stock sample, with control variables					
1	0.00	−0.02	0.01	−0.01	0.03
2	−0.01	−0.02	0.02	−0.07	0.03
3	0.02	−0.06	0.03	0.02	0.10**
4	0.02	−0.01	0.04*	−0.03	0.05
Flow sample, with control variables					
1	−0.05	−0.14*	0.00	−0.14**	0.02
2	−0.03	0.04	0.01	−0.19	0.02
3	0.01	0.00	−0.04	−0.05	0.18
4	−0.02	0.01	−0.03	−0.04	0.09
Stock sample, without control variables					
1	0.00	−0.02	0.01	−0.03	0.02
2	−0.01	−0.02	0.02	−0.11	0.02
3	0.02	−0.06	0.04	0.03	0.10**
4	0.03	0.00	0.05**	−0.04	0.06
Flow sample, without control variables					
1	−0.04	−0.16*	0.00	−0.09	0.04
2	−0.02	0.01	0.01	−0.13	0.04
3	0.01	0.00	−0.03	0.00	0.18
4	−0.01	−0.01	−0.02	0.00	0.09

Note: Instrumental variable estimates of the effect of downloading the predictions on compliance, where compliance is measured in the long term (Version B). (Geneva is not included since predictions were hardly ever downloaded.) Standard errors are clustered by caseworkers. Significance at the 1%, 5% and 10% level is marked with ***, ** and *, respectively.

References

BLACK, D., J. SMITH, M. BERGER, and B. NOEL (2003), "Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System", *American Economic Review* 93, pp. 1313–1327.

Black, D., J. Smith, M. Plesca, and S. Plourde (2002), "Profiling UI claimants to Allocate Employment Services: Evidence and Recommendations for the States", Final report to the US Department of Labour.

Behncke, S., M. Frölich, and M. Lechner (2007), „Pilotprojekt Statistisch assistierte Programmselektion (SAPS)", Swiss State Secretariat for Economic Affairs (seco), www.seco.admin.ch/dokumentation/publikation/00004/00005/index.html?lang=de.

Bell, S., and L. Orr (2002), "Screening (and Creaming?) Applicants to Job Training Programs: The AFDC Homemaker-Home Health Aide Demonstrations", *Labour Economics* 9, pp. 279–301.

Berger, M., D. Black, and J. Smith (2001), "Evaluating Profiling as a Means of Allocating Government Services", in: *Econometric Evaluation of Labour Market Policies*, M. Lechner and F. Pfeiffer (eds), Physica/Springer, Heidelberg, pp. 59–84.

Brownell, K. and T. Wadden (1991), "The Heterogeneity of Obesity: Fitting Treatments to Individuals", *Behavior Therapy*, 22, pp. 153–177.

Bryson, A., and D. Kasparova (2003), "Profiling Benefit Claimants in Britain: A Feasibility Study", Department for Work and Pensions, Research Report no. 196.

Colpitts, T. (2002), "Targeting Reemployment Services in Canada: The Service and Outcome Measurement System (SOMS) Experience", in: Eberts, R. W., D. J. O'Leary, S. Wandner (eds), *Targeting Employment Services*, Kalamazoo, Michigan: W.E. Upjohn Institute, pp. 283–301.

Dehejia, R.H. (2005), "Program Evaluation as a Decision Problem", *Journal of Econometrics* 125, pp. 141–173.

De Koning, J. (1999), "The Chance-Meter: Measuring the Individual Chance of Long-Term Unemployment", Paper presented at the Conference 'Current Developments in the Evaluation of Employment Policies', Barcelona, 19–20 July 1999, Division of Labour and Education, Rotterdam.

Dickinson, K., P. Decker, and S. Kreutzer (2002), "Evaluation of WPRS sytems", in: Eberts, R. W., D. J. O'Leary, S. Wandner (eds), *Targeting Employment Services*, Kalamazoo, Michigan: W. E. Upjohn Institute, pp. 61–81.

Eberts, R. (2002), "Design, Implementation and Evaluation of the Work First Profiling Pilot Project", Report for the US Department of Labor. http://wdr.doleta.gov/research/keyword.cfm?fuseaction=dsp_resultDetails&pub_id=2287&bas_option=Keywords&start=1&usrt=4&stype=basic&sv=1&criteria=work%20first .

EBERTS, R., and C. O'LEARY (2002), "A Frontline Decision Support System for Georgia Career Centres", W. E. Upjohn Institute Staff Working Paper no. 02-84, Kalamazoo, Michigan.

EBERTS, R., C. O'LEARY and S. WANDNER (2002), *Targeting Employment Services*, Kalamazoo, Michigan: W. E. Upjohn Institute.

EGGER, M., and C. LENZ (2006), "Wirkungsevaluation der öffentlichen Arbeitsvermittlung – Evaluationsbericht", SECO Publikation Arbeitsmarktpolitik No 18 (10.2006), Swiss State Secretariat for Economic Affairs (seco), www.seco. admin.ch/dokumentation/publikation/00004/00005/index.html?lang=de. French version: www.seco.admin.ch/dokumentation/publikation/00004/ 00005/index.html?lang=fr.

FRASER, N. (2000), "How Strong is the Case for Targeting Active Labour Market Policies", *International Journal of Manpower* 20, pp. 151–164.

FREDRIKSSON, P., and P. JOHANSSON (2003), "Program Evaluation and Random Program Starts", IFAU Discussion Paper 2003:1.

FRÖLICH (2006), "Statistical Treatment Choice: An Application to Active Labour Market Programmes", IZA Discussion Paper 2187.

FRÖLICH (2008), "Statistical Treatment Choice: An Application to Active Labour Market Programmes", forthcoming in the *Journal of the American Statistical Association*.

FRÖLICH, M., M. LECHNER, S. BEHNCKE, S. HAMMER, N. SCHMIDT, S. MENEGALE, A. LEHMANN, and R. ITEN (2007), "Einfluss der Rav auf die Wiedereingliederung von Stellensuchenden", SECO Publikation Arbeitsmarktpolitik no. 20 (1.2007), Swiss State Secretariat for Economic Affairs (seco), www.seco.admin.ch/dokumentation/publikation/00004/00005/index.html? lang=de.

FRÖLICH, M., M. LECHNER, and H. STEIGER (2003), "Statistically Assisted Programme Selection – International Experiences and Potential Benefits for Switzerland", *Swiss Journal of Economics and Statistics* 139, pp. 311–331.

HORRACE, W., and P. SCHMIDT (2000), "Multiple Comparisons with the Best, with Economic Applications", *Journal of Applied Econometrics* 15, pp. 1–26.

KEUM, J. (2001), "Review of the Effectiveness of Public Employment Services", in: Park, F., Y. Park, G. Betcherman, A. Dar (eds), *Labor Market Reforms in Korea: Policy Options for the Future*, Korea Labor Institute.

KREUTER, M., and V. STRECHER (1996), "Do Tailored Behavior Change Messages Enhance the Effectiveness of Health Risk Appraisals?: Results from a Randomized Trial", *Health Education Research* 11, pp. 97–105.

Lalive, R., J. van Ours, and J. Zweimüller (2005), "The Effect Of Benefit Sanctions On The Duration Of Unemployment", *Journal of the European Economic Association* 3, pp. 1386–1417.

Lechner, M., and J. Smith (2007), "What is the Value Added by Caseworkers?", *Labour Economics* 14, pp. 135–151.

Manski, C. (2000), "Identification Problems and Decisions under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice", *Journal of Econometrics* 95, pp. 415–442.

Manski, C. (2004), "Statistical Treatment Rules for Heterogeneous Populations", *Econometrica* 72, pp. 1221–1246.

Manski, C. (2006), "Search Profiling With Partial Knowledge of Deterrence", *The Economic Journal* 116, pp. F385–F401

Manski, C. (2007), "Minimax-Regret Treatment Choice with Missing Outcome Data", *Journal of Econometrics* 139, pp. 105–115.

Moffitt, R. (2006), "Welfare Work Requirements with Paternalistic Government Preferences", *The Economic Journal* 116, pp. F441–F458.

Moisala, J., I. Suoniemi, and R. Uusitalo (2006), "Työttömien tilastollinen profilointi", http://www.mol.fi/mol/fi/ 99_pdf/ fi/ 06_tyoministerio/ 06_julkaisut/06_tutkimus/tpt316.pdf.

Murphy, S. (2003), "Optimal Dynamic Treatment Regimes", *Journal of Royal Statistical Society,* Series B 65, pp. 331–366.

Murphy, S., K. Lynch, D. Oslin, J. McKay, and T. TenHave (2007), "Developing Adaptive Treatment Strategies in Substance Abuse Research", *Drug and Alcohol Dependence*, Supplement 2, 88S, pp. S24–S30.

OECD (1998), "Early Identification of Jobseekers at Risk of Long-Term Unemployment: The Role of Profiling", OECD Proceedings, Paris.

O'Leary, D., P. Decker, and S. Wandner (2005), "Cost Effectiveness of Targeted Reemployment Bonuses", *Journal of Human Resources* 40, pp. 270–279.

Plesca, M., and J. Smith (2005), "Rules versus Discretion in Social Programs: Empirical Evidence on Profiling in Employment and Training Programs", unpublished, University of Michigan.

Persico, N., and P. Todd (2005), "Passenger Profiling, Imperfect Screening, and Airport Security*, American Economic Review, Papers and Proceedings* 95, pp. 127–131.

Persico, N., and P. Todd (2006), "Generalising the Hit Rates Test for Racial Bias in Law Enforcement, With an Application to Vehicle Searches in Wichita", *The Economic Journal* 116, pp. F351–F367.

Rudolph, H., and M. Müntnich (2001), "Profiling zur Vermeidung von Langzeitarbeitslosigkeit – erste Ergebnisse aus einem Modellprojekt", *MittAB*, pp. 530–553.

Rush, A. (2005), "Algorithm-Guided Treatment in Depression: TMAP and STAR*D", in M. Bauer, A. Berghofer and M. Adli (eds), *Therapieresistente Depressionen – Aktueller Wissensstand und Leitlinien für die Behandlung in Klinik und Praxis*, Springer, Heidelberg.

Smith, J. (2002), "Comments on Chapter 10", in: Eberts, R.W., D.J. O'Leary, S. Wandner (eds), *Targeting Employment Services*, Kalamazoo, Michigan: W.E. Upjohn Institute, 2002, pp. 303–308.

Staghøj, J., M. Svarer, and M. Rosholm (2007), "A Statistical Programme Assignment Model", mimeo.

Stephan, G., S. Rässler, and T. Schewe (2006), "Das TrEffeR-Projekt der Bundesagentur für Arbeit. Die Wirkung von Maßnahmen aktiver Arbeitsmarktpolitik", *Zeitschrift für Arbeitsmarktforschung*, 39, pp. 447–465.

U.S. Department of Labor, Employment and Training Administration (1999), "Evaluation of Worker Profiling and Reemployment Services Policy Workgroup: Final Report and Recommendations", Washington D.C.

Velicer, W., and J. Prochaska (1999), "An Expert System Intervention for Smoking Cessation", *Patient Education and Counselling* 36, pp. 119–129.

Wandner, S. (2002), "Targeting Employment Services under the Workforce Investment Act", in: Eberts, R.W., D.J. O'Leary, S. Wandner (eds.), *Targeting Employment Services*, Kalamazoo, Michigan: W.E. Upjohn Institute, 2002, 1–25.

Wunsch, C. (2007), "Optimal Use of Labour Market Policies", Discussion Paper 2007-26, Department of Economics, University of St.Gallen.

SUMMARY

We evaluate a randomized experiment of a statistical support system developed to assist caseworkers in Swiss employment offices in choosing appropriate active labour market programmes for their unemployed clients. This statistical support system predicted the labour market outcome for each programme and thereby suggested an 'optimal' labour market programme for each unemployed person. The support system was piloted in several employment offices. In those pilot offices, half of the caseworkers used the system and the other half acted as control group. The allocation of the caseworkers to treatment and control group was random. The experiment was designed such that caseworkers retained full

discretion about the choice of active labour market programmes, and the evaluation results showed that caseworkers largely did not follow the statistical support system. This indicates that stronger incentives are needed for caseworkers to comply with statistical profiling and targeting systems.