Disentangling Treatment Effects of Active Labor Market Policies:
The Role of Labor Force Status Sequences[*]

Jochen **Kluve**[a**]

Hartmut **Lehmann**[b]

Christoph M. **Schmidt**[c]

This version: December 2007

Abstract. This paper estimates treatment effects of two active labor market policies – a training program and a wage subsidy scheme – on participants' employment probabilities. The analysis is based on unique data from the 18th wave of the Polish Labor Force Survey containing detailed and extensive individual labor force status histories. We discuss two variants of an exact covariate matching procedure adapted to the specific nature of the data. Our study confirms and reinforces a point raised in recent research (Heckman and Smith 1999, 2004), that pre-treatment labor force status dynamics play a decisive role in determining program participation. We implement a conditional difference-in-differences estimator of treatment effects based on these individual trinomial sequences of pre-treatment labor market status. The estimator employs a "moving window" technique that nicely controls for changes in the macroeconomic environment over time. Our findings suggest that training raises individual employment probability, while wage subsidies display negative treatment effects for participants in the Polish case.

Keywords: Active Labor Market Policy, exact matching, moving window, determinants of program participation.
JEL: C49, J68

---

[a] RWI Essen, and IZA Bonn.
[b] DARRT, University of Bologna, IZA Bonn, and CERT, Heriot-Watt University, Edinburgh.
[c] RWI Essen, Ruhr-Universität Bochum, CEPR London, and IZA Bonn.

** Corresponding author: kluve@rwi-essen.de.

1. Introduction

Over the last decade there has been much interest by labor economists in the evaluation of so-called Active Labor Market Policy (ALMP), i.e. policy measures such as training programs, wage subsidy schemes, or direct job provision in the public sector. These measures, generally, aim at increasing the employment probability and/or the earnings performance of program participants. In the US, experiences with both the implementation and evaluation of such programs date back well into the 1960s (Heckman, LaLonde and Smith 1999). In Europe, where unemployment had remained comparatively low until the 1980s, or even until the 1990s in some countries, running such programs and evaluating them is a rather recent phenomenon. Nevertheless, most countries in Western Europe have now utilized active labor market measures for many years, and have done so with substantial financial input in terms of fraction of GDP spent on these measures (see e.g. OECD 2000, 2004). Also the evaluation practice, while still lagging behind the US "evaluation culture" to some extent, has attained increasing interest – and funding – by European policy makers, both in individual countries and from the European Commission.[1]

Looking further east we observe that – after the breakdown of the socialist regimes and the beginning of the "transition process" around 1990 – Eastern European countries were confronted with the task of redesigning their welfare system. Suddenly facing substantial open unemployment, schemes for passive and active support of unemployed individuals had to be set up from scratch. Frequently this led to transition countries' adoption of Western schemes, often without much knowledge about their efficiency. Poland, too, implemented a system of unemployment benefit support, accompanied by a set of Active Labor Market Policies. Specifically, two significant active programs for the unemployed implemented in Poland are (i) a training program, and (ii) "Intervention Works", a wage subsidy scheme.

While the importance of such programs – as expressed in government spending in terms of GDP share – has declined in Poland over recent years, a few studies of Polish labor market dynamics and Active Labor Market Policy evaluation were undertaken in the late 1990s (Góra and Schmidt 1998, Puhani 1998). In this paper, we will build on earlier work on program evaluation in Poland (Kluve, Lehmann and Schmidt 1999) and provide

---

[1] Kluve and Schmidt (2002) and Kluve (2006) give accounts of the European experience with Active Labor Market Policies and contrast this experience with the evidence from the US.

an in-depth investigation of the Polish experience with Active Labor Market Policy in the mid-1990s, specifically the years 1992 to 1996.

There is a set of features that make this study particularly interesting. First, we use data from the 18th wave of the Polish Labor Force Survey (PLFS). The data were collected in August 1996, and contained a supplementary questionnaire on past labor market experience of respondents. This supplement generated a unique set of individual labor market histories dating from January 1992 until August 1996, and comprising a person's labor force status for every single month. The monthly labor force status captures employment, unemployment, inactivity, and participation in an active labor market program, as well as a set of other states, such as caring for a child etc., that are of minor interest to our study.

Second, the evaluation is set against the background of a country in the early years of transition. This implies a rapidly changing macroeconomic environment, making it indispensable to develop a treatment effect estimator that can account for these changes in an appropriate manner. Third, in addition to estimating treatment effects on the basis of individual labor force status histories, we can use the detailed monthly data to investigate further how important such labor market histories are in fact for determining participation in the program, and hence the evaluation approach. Recent research (Heckman and Smith 2004, complementing Heckman and Smith 1999) suggests that labor force status dynamics play the central role in driving participation dynamics. We will reinforce this point made by Heckman and Smith on the basis of a different data set, different active labor market programs implemented in a different country, in an entirely different context.

The core part of our analysis is the development of a matching estimator based on individual pre-treatment labor force status sequences. This creates a "moving window" structure that allows for individually flexible entry into and exit out of the program, hence conditioning on covariates and labor force status sequences at exactly the month of program start, and comparing outcomes at exactly the month of program termination. Clearly, this procedure not only increases comparability of treated and comparison units, but also nicely controls for changes in the macroeconomic environment. Our approach is delineated using two matched samples, for two active policy measures – (i) Training and (ii) Intervention Works, respectively, created from the reservoir of program participants and a comparison group consisting of those untreated individuals that were unemployed at least

once over the sample period.[2] For a sample (A) the comparison group is matched on a set of covariates, in particular taking into account the local labor market context, a variable whose importance in program evaluation is e.g. pointed out in Heckman, Ichimura and Todd (1997). In addition, for a sample (B) matches are formed on both covariates and four-quarter individual pre-treatment labor market histories, in the spirit of Card and Sullivan (1988).

The paper is organized as follows. Section 2 describes the institutional set-up and expenditure patterns of ALMP in Poland during our sampling period (1992-1996). The third section delineates our data and the matching approach to program evaluation. In section 4 we discuss the matched samples, focusing on the timing of interventions and the role of pre-treatment labor force status histories. Section 5 presents our estimation strategy and estimation results. Section 6 concludes.

2. Active Labor Market Policies in Poland

Expenditures on labor market policies by the Polish government disbursed by the Labor Fund[3] have been rather limited in international perspective. As **Table 1** demonstrates, in the years 1990 to 1996 Poland spent roughly 2 percent of GDP on passive and active labor market policies, compared to about 3 percent of GDP spent by most OECD countries, many of which had far lower unemployment rates than Poland. As we can also see from Table 1, apart from 1990 when registered unemployment was relatively low, expenditures on ALMP are only a small fraction of total expenditures on the unemployed, never exceeding one eighth of total expenditures. Kubiak (2005) makes convincingly the point that throughout the 1990s outlays on ALMP in real terms are "crowded out" by the payment of unemployment benefits: as the number of unemployed rises, these outlays fall and vice versa.

In addition to training and intervention works, the two measures evaluated in this paper, labor offices also funded the following schemes: public works, loans for entrepreneurial activity, and a measure to integrate unemployed school leavers into the labor market. **Table 2** shows, however, that training and intervention works consumed more than fifty percent of all expenditures on ALMP measures in the years 1991 to 1996.

---

[2] Eligibility of program participation was confined to people registered with the local labor office as unemployed.

[3] The Labor Fund is a special fund that is financed by a 3 percent payroll tax to be paid by employers and by donations from the central budget (Golinowska 1999). All expenditures in connection with training and intervention works are paid from this fund.

Between 1992 and 1996 the two measures have relatively large stocks of participants (cf. **Table 3**), while the number of placements of unemployed in public works slots is small in the first two years and only thereafter growing. The measure to integrate unemployed school leavers becomes an important element of Polish ALMP only after 1996.[4] So, during the period for which we have individual labor market history data (cf. subsequent sections) only training and intervention works were measures with large enough stocks of participants that allow a rigorous microeconometric analysis of their efficacy.

2.1 Institutional issues and unemployment benefit regulations for the general pool of unemployed

Even before the beginning of economic reform in January 1990 the Polish government established a legal framework that defined the scope and nature of passive and active labor market policies (ALMP). The Law on Employment passed in December 1989[5] had several modifications and extensions over the years. For the discussion of the two ALMP measures during the reported period in our paper (1992 to 1996) two new laws modifying and extending the original law are relevant, namely the Law on Employment and Unemployment from 1991[6] and the Law on Employment and Measures to Combat Unemployment from 1994.[7] The great bulk of the regulations of the 1994 law are still valid today, i.e. there has been substantial inertia in labor market legislation. Before discussing the two ALMP measures in detail, we will briefly describe the unemployment benefit regulations and their evolution over the relevant period for those unemployed who did not participate in a program.

The provisions for unemployment benefits for the general pool of unemployed went through several stages of major revisions. The original law of 1989 foresaw open-ended as well as earnings-related benefits and did not impose any previous employment requirement. Unemployed persons who registered with the local employment office were

---

[4] In 1990 loans for entrepreneurial activity were the predominant ALMP measure, probably because this measure required little administrative infrastructure that still needed to be built at that time. Since 1991 these loans have been marginal, while the measure to integrate unemployed school leavers did not exist in the first years but became as important as intervention works and public works towards the end of the nineties. For a discussion of these trends, see Kubiak (2005).

[5] The law was published under the Polish name "Ustawa o Zatrudnieniu" in the Polish Government's official bulletin "Dziennik Ustaw" in December 1989 (number 75, position 446).

[6] The "Ustawa o Zatrudnienie i Bezrobociu" was published in the Dziennik Ustaw in October 1991 (number 106, position 457).

[7] The "Ustawa o Zatrudnieniu i Przeciwdzialaniu Bezrobocia" was published in the Dziennik Ustaw in January 1995 (number 1, position 1).

entitled to such benefits seven days after registration. They received 70 percent of the net wage at their last employment for the first three months of unemployment, 50 percent for the next six months and 40 percent when their spell was longer than nine months. Persons who had never worked received the minimum wage. Local labor offices applied ILO criteria when determining whether a registered person was unemployed.

The amendments incorporated in the 1991 law foresaw a previous employment requirement: persons had to have worked over the last year before registration for at least 180 days to be entitled to benefits. While the earnings-related nature of benefits was still kept with this new law, benefit payments were now limited in general to 12 months. There were, however, some important exceptions to this last rule. Workers with at least 25 years of experience if female and 30 years if male were entitled to 18 months of benefits if they became unemployed. Unemployed who were within two years of retirement and who had the above mentioned long work experience could receive benefits until they retired. Since under central planning all able bodied employed persons had continuous employment relationships over their working life, using age and gender as matching criteria ensures virtually in all cases that unemployment benefit rules are identical for treated persons and their controls.

The last law relevant for our paper, which took effect on the 1st of January 1995, introduced two major changes. It introduced a flat rate of benefits, amounting to 36 percent of the average wage in the economy and it allowed for a variation of the length of benefits by region. The latter regulation foresaw that in voivodships (regions) in which the unemployment rate was 50 percent below (above) the national average the entitlement *could* be decreased (increased) to 6 (18) months. While, subsequently, this provision was rarely used, by matching on voivodship we are not only trying to control for local labor market conditions but also ensuring that in the few cases when this provision might apply both treated and controls have the same potential length of benefits.

2.2 Intervention works (prace interwencyjne)

Intervention works is essentially a measure of subsidized employment. In the reported period, firms could create subsidized jobs if they had not laid off in the previous year more than 10 percent of their workforce or if they were not in the state of liquidation or bankruptcy. Also, intervention works jobs could only be created if they did not crowd out regular non-subsidized jobs in a sector. Until 1994 the maximum duration of these

subsidized jobs could be six months; after the law was amended in December 1994 the maximum duration was stipulated to be twelve months. Local labor offices used resources from the Labor Fund to reimburse a part of the costs of a subsidized job for a period of six months (until 1994) or up to twelve months (after 1994). Before the law was amended the Labor Fund paid wage costs amounting to unemployment benefits and social security contributions if the participant was entitled to benefits. If the person had no benefit entitlements, wage costs as large as the minimum unemployment benefits were reimbursed as well as the paid social security contributions. After 1994, the reimbursement policy was slightly more complicated, as the costs of jobs lasting up to six months were reimbursed as before, while with jobs lasting between six and twelve months the Labor Fund paid all wage costs every second month, as long as these costs did not exceed the minimum earnings at these jobs in the firm and the corresponding social security contributions. If a worker was retained on the job for an indefinite period after the end of the subsidized spell, the firm received a premium of one monthly national average remuneration (until 1994) or of 150 percent of one monthly national average remuneration (after 1994).

The unemployed participating in an intervention works job received the going wage and fringe benefits for the assumed position at the specific firm. An unemployed person who did not find regular employment after the end of the subsidized job was entitled to twelve months of unemployment benefits. Before the law was amended at the end of 1994 this entitlement was extended to all unemployed on an intervention works job, even if this job had lasted less than 180 days. With the amendment of the law, this stipulation was removed and unemployed were only entitled to benefits after a subsidized job spell if this spell had lasted at least 180 days, reflecting the general rule that benefits entitlement was only given to those who in the previous year had worked at least 180 days. When entitled to a new round of benefits, a person was subject to the same rules and regulations as a person from the general pool of unemployed.

The unemployed did not select themselves into the intervention works jobs but were selected by officials of local employment offices. In the years 1992 to 1996, two decrees by the ministry of labor and social policy gave some guidance on whom to select into the program.[8] The decree at the end of 1991 foresaw that an employment office had to take the age and the health status into consideration when choosing participants, as well as

---

[8] The decrees of 17th December 1991 and of 21st of March 1995 on the question of the organization of intervention works and public works.

the benefits accruing to society. Apart from these rather vague stipulations, employment offices were directed to choose unemployed school leavers, persons with unemployment spells of more than six months, unemployed with the appropriate skills and persons with prospects of continuing employment after the termination of the subsidized job spell. The second decree, passed in March of 1995, foresaw the selection of persons with unemployment spells of more than twelve months, of persons whose spouse was also unemployed and of unemployed single parents. The two stipulations about the selection of unemployed with the appropriate skills and persons with prospects of continuing employment after the termination of the subsidized job spell appeared in both decrees, as well as the vague notion that participation in the intervention works scheme had to be beneficial to society.

2.3 Training (Szkolenie zawodowe)

In the reported period, the training scheme involved retraining and further training with the stated aim of attenuating skill mismatch. It was initiated and financed by local employment offices from resources made available by the Labor Fund. Unemployed workers but also workers who were still employed but faced redundancy could participate in training measures. The training for the unemployed took place predominantly in private training institutions while employed trainees did undergo retraining or further training at the firm that had approached the local employment office for financial support of these training measures. According to the law, the training should not exceed six months, however, under exceptional circumstances a training period of up to twelve months was considered admissible. Local employment offices financed up to fifty percent of the training costs of still employed workers who faced redundancy.

In the years 1992 to 1996, there were three types of participants in the training program. First, unemployed who were selected by local employment offices for training. Second, unemployed who themselves chose a retraining or further training scheme, with the local employment office being prompted by law to finance this training if there was a high probability that after the training the unemployed could enter a regular job. The third category were still employed workers who faced redundancy for reasons that had to do with the economic situation of the firm. The law passed at the end of 1994 foresaw a fourth type of participants. Firms with more than fifty employees could apply for training subsidies in order to retrain workers who were threatened by redundancy at their existing

jobs and who had prospects of being employing in different positions at the same firm for a period of at least twelve months after the end of the training. These training subsidies amounted to up to fifty percent of the training costs.

The law passed in October 1991 established detailed rules about the remuneration of unemployed training participants while on a training course. We report here the rules for those types of unemployed who made up the bulk of training participants. Those unemployed who lost the ability to perform their job because of a work accident or because of professional illness received 100 percent of their earnings in their last job, but at least 40 percent and not more than 110 percent of average earnings in the economy. For unemployed made redundant this remuneration while training amounted to 80 percent of earnings in the last job, with the above mentioned lower and upper bounds related to average national earnings. Unemployed school leavers participating in training received 115 percent of the unemployment benefits to which they were entitled, while those unemployed training participants who had never worked before did not receive any income support.

The law of 1991 was very generous for trainees as far as benefit entitlements were concerned. An unemployed person undergoing training was entitled to twelve months of benefits if the local employment office was not able to mediate a regular job after the unemployed had finished the training or was not putting the trainee on an intervention works or public works scheme. The law of 1994, taking effect on January 1st 1995, eliminated this generosity. The trainee could receive only one month of benefit payment after the end of the training and only in the case that s/he was entitled to benefits during the last month before taking up the training.

3. Data and Methods

3.1 The Data

We employ data from the 18th wave of the Polish Labour Force Survey (PLFS) as of August 1996. The PLFS is a quarterly rotating panel introduced in May 1992. The distinct feature of the August 1996 wave is a supplementary questionnaire containing retrospective questions on individual labor market behavior. Specifically, the questionnaire allows constructing individual labor market histories on the basis of labor force status in every single month. Possible states are employed, unemployed, inactive, program participation, etc. (see below). The individual histories cover the 56-month-period from January 1992 to

August 1996.

Our evaluation of the Training and Intervention Works programs is based on considering (a) pre-treatment labor force status information over a period of 4 quarters, i.e. 12 months, and (b) post-treatment employment outcomes over a period of 3 quarters, i.e. 9 months. Given an overall sampling period from January 1992 until August 1996, we therefore focus on individuals whose treatment started after December 1992 and ended before December 1995. The analysis takes into account all individuals who experienced at least one spell of unemployment during the observation period. For both treated units and potential comparison units this ensures consideration of individuals potentially eligible for participation in ALMP measures offered by the labor offices. We discuss sample composition in more detail in section 4.1.

In order to be able to handle such rich data, we had to condense the information contained in the individual labor market histories. Monthly entries comprise, for instance, states such as "employed", "unemployed", "receiving unemployment benefits", "maternal leave", etc. Furthermore, individual histories indicate whether and when an individual took part in an ALMP course. We compress the 30 possible monthly states occurring in the data into the three labor market states "employed" (henceforth denoted "1"), "unemployed" (denoted "2"), and "out-of-the-labor-force" (denoted "0"). Information on treatment participation is stored separately. Kluve et al. (1999) give a more detailed account of data transformation and adaptation. The resulting structure of individual spells for treatment and potential comparisons will be illustrated further in section 4.2.

In the estimation of individual treatment effects we consider two distinct measures of Polish ALMP, (i) Training and (ii) Intervention Works. (i) Training is meant to enhance, or at least sustain, individual human capital. The Polish Training measure predominantly entails training off-the-job[9] whose final aim is raising a non-employed person's probability of re-employment in a regular job. A minor component of this measure consists in preparing persons still employed in jobs threatened by redundancy for viable jobs in the same firm or elsewhere. Hence, the overriding aim is to attenuate skill mismatch. (ii) Wage subsidy schemes like the Polish Intervention Works also have a human capital enhancing or -preserving aspect. The enhancement or preservation of a person's human capital takes

---

[9] Since off-the-job training entails training of persons who are either unemployed or out-of-the-labor-force when training started, we can use the pre-treatment labor force status sequences to estimate the fraction of training participants engaged in this type of training. Before the law was amended in December 1994, 73 percent of participants engaged in off-the-job training, while after December 1994 this fraction amounted to 69 percent.

place on-the-job. This human capital component of the program is thought to increase the chances of a participant to find regular, non-subsidized employment at the same firm or elsewhere after the end of the program. In addition, if there is asymmetric information about the productivity of potential employees, wage subsidy schemes are designed to facilitate temporary job matches that might translate into regular and lasting matches at the same firm once the subsidy ends.

A crucial feature of ALMP regulation in the reported period, however, was that participation in Intervention Works was considered by the law like any other employment spell, hence in essence entitling individuals to a new round of benefit receipt. Taking part in a Polish training measure, on the other hand, did not entitle the large majority of participants to unemployment benefits.[10]

3.2 Matching Methods

Program evaluation aims at estimating causal effects of treatments, i.e. changes in the outcome variable of interest that are due to participation in the treatment. The application of matching methods for treatment effect estimation has become quite popular over recent years, and several variants of matching estimators are now routinely applied.[11] The causal model underlying this approach has become known as the "Potential Outcome Model" and is based on work by Neyman (1923 [1990]), Fisher (1935) and Rubin (1974, 1977; see also Holland 1986 for discussion). The model formalizes the idea that, in order to infer a causal effect of the treatment on the outcome variable, it is necessary to identify the counterfactual, i.e. what would have happened to the treatment group if it had not been exposed to treatment? Then the causal effect of treatment is given by the difference between the factual (=exposed to treatment) and counterfactual (=not exposed to treatment) outcomes for the participant population.

Let the binary variable $D_i \in \{0,1\}$ indicate the treatment received, i.e. $D_i = 1$ if the

---

[10] As discussed in section 2, before the law was changed in December 1994 participants had their benefit entitlement renewed no matter what the duration of the intervention work job, while starting with January 1995 the subsidized job had to have lasted at least 180 days, i.e. six months. Since roughly three quarters of intervention works participants in our sample started the scheme before December 1994 and since the duration of two thirds of those intervention works jobs started in January 1995 or later amounted to at least six months, the vast majority of participants was clearly entitled to a new round of benefits. Roughly two thirds of participants in training started the measure after December 1994, when it had become virtually impossible for trainees to receive unemployment benefits.

[11] Much research has been conducted in labor economics and econometrics on the practical and theoretical properties of matching estimators. See, for instance, the debate between Dehejia/Wahba and Smith/Todd (Dehejia and Wahba 1999, 2002, Dehejia 2005, Smith and Todd 2005a, 2005b) and a recent symposium in the Review of Economics and Statistics (2004, Vol. 86, No. 1, pp. 1-194).

unemployed individual participates in the program. For each person we observe the treatment that she actually received, and the outcome associated with this treatment, i.e.

$$Y_i = Y_{0i} \quad if \ D_i = 0,$$
$$Y_i = Y_{1i} \quad if \ D_i = 1,$$

where Y_i captures post-treatment outcomes of the variable of interest, here specifically individual labor market status ("1" when the worker is employed and "0" otherwise). Thus, the unit level causal effect given by $\Delta_i = Y_{1i} - Y_{0i}$ is never directly observable. The essential conceptual point is that nonetheless each individual has two possible outcomes associated with herself, where one realization of the outcome variable can actually be observed for each individual, and the other one is a counterfactual outcome.

Since individual-level effects cannot be observed, the estimand of interest should be a measure that summarizes individual gains from treatment appropriately. One parameter that has received particular interest in the program evaluation literature is the average treatment effect for the treated population (ATET),

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

where the expectations operator E(.) denotes population averages. The parameter is generally not identified from observational data: Whereas the first of the population averages in the ATET parameter can be identified for the treatment group subsample, the counterfactual expectation $E(Y_0 \mid D = 1)$ is not identifiable without invoking further assumptions, since the outcome under no-treatment is not observed for the treated population. This is precisely the counterfactual of interest: What outcome would the treated units have realized if they had not been exposed to the treatment?

If treatment is not randomly assigned, matching intends to mimic a randomized experiment *ex post*. This strategy is feasible if there is only "overt bias" (Rosenbaum 1995), i.e. treatment and comparison group differ prior to treatment only in observable variables that matter for the outcome under study. Let X denote the vector of observed pre-treatment variables, or covariates. Then the concept of "selection on observables" is formalized in the following *identifying assumption*: The assignment mechanism D is independent of the potential outcomes Y_0 and Y_1 conditional on X (Rubin 1974, 1977). This assumption is commonly referred to as *unconfoundedness* (Imbens 2004). By the unconfoundedness assumption it is possible to replace the no-treatment outcome for the treated population with the no-treatment outcome of the non-treated, i.e. comparison,

population:

$$E(\Delta \mid X, D = 1) = E(Y_1 \mid X, D = 1) - E(Y_0 \mid X, D = 1)$$
$$(2) \qquad = E(Y_1 \mid X, D = 1) - E(Y_0 \mid X, D = 0).$$

This covariate-adjusted ATET is identified from observable data.

4. Analyzing Matched Samples

4.1 Composition of Matched Samples

For each of the two active labor market measures under scrutiny – Training and Intervention Works – we analyze treatment effects and illustrate the role of labor force status sequences using two samples. As the reservoir from which comparison units are drawn, we consider those observations in the PLFS that have at least one spell of unemployment over the sampling period January 1992 to August 1996, and whose timing structure would potentially allow them to find a matching partner among the treated workers. The two matched samples are then defined as follows.

> **Sample A**: A comparison unit is matched to a treated unit, if his or her labor market history is observed without substantial gaps for 12 months preceding the start of treatment **and** for 9 months succeeding the end of treatment, and if he or she is identical in observable covariates age, gender, education, marital status, and region.

> **Sample B**: A comparison unit is matched to a treated unit, if the requirements for sample (A) are met, and if he or she displays an **identical** 4-quarter (12-month) pre-treatment labor market history at the exact same point in time as the treated unit.[12]

The matching algorithm used to construct samples (A) and (B) applies exact covariate matching. For both samples, if a treated individual finds any matching partner among the potential comparisons, this observation is retained. The algorithm allows for an oversampling procedure, i.e. a treated unit may be assigned more than one comparison unit. This procedure reflects our conviction that **timing** is the pivotal aspect of comparison group construction in a transition economy (see also next subsection). Under requirement

---

[12] We consider 6 age categories, 3 education categories, gender, marital status, and 49 regions, resulting in 3528 different cells for sample (A). Including a 4-quarter sequence of a trinomial labor market outcome variable (cf. section 4.2) increases the number of cells to $3528*3^4=285,768$ cells for sample (B).

(B) fewer treated units find matching partners than under (A), and the number of matched comparison units is smaller. Thus, algorithm (B) proceeds with replacement: some comparison units are matched to more than one treated individual.

Section 2 above has detailed that the selection into programs followed certain stipulations, which were broadly specified rather than specific. They did, however, clearly put the main emphasis on individuals' previous labor market experience and human capital acquisition. We thus believe that in the given text the unconfoundedness assumption is justified, basing our analysis on a set of core covariates – including education and age, which also played a role in selecting participants – in combination with very detailed labor force status histories. Moreover, we take into account the importance of local labor market conditions (cf. Heckman, Ichimura and Todd 1997) by means of an indicator variable for the 49 Polish regions ("voivodships").

Table 4 presents sample sizes and covariate means for samples (A) and (B). We observe that there is a reduction in the number of treated units who find matching partners when moving from the reservoir to sample (B) of approximately one quarter for Training, and less than one quarter for Intervention Works. Due to matching-with-replacement, samples (B) contain comparison units matched to more than one treated unit. With less than one percent, the number is very low for Training, and with approximately one tenth it is also fairly low for Intervention Works. Table 4 also shows that Training participants on average are better educated, somewhat younger and more likely to be female than Intervention Works participants. Clearly, while the reservoir displays imbalances in the covariates, matching in samples (A) and (B) by definition produces balance, since it conditions on identicalness in observed characteristics age, education, sex, marital status, and local labor market (region).

4.2 The timing of interventions

In sample (B) we require treated and matched comparison units to display an identical pre-treatment history. To achieve comparability across the two samples (A) and (B), we impose the requirement on sample (A) that we observe any history at all in the year preceding treatment, although the precise information *what* history was experienced is not used in matching[13]. Moreover, to allow an assessment of post-treatment labor market performance, we require treated units and comparison units in all samples to have a complete post-

---

[13] This requirement was also used to define the reservoir.

treatment sequence of labor force status in the nine months after treatment. Monthly employment information is condensed into a sequence of three quarters of a multinomial outcome variable (0,1,2) denoting labor force status (out-of-the-labor force, employed, unemployed). The state "employed" overrules the other two states, i.e. if a person is employed in at least one month of a quarter, we consider her employed during that quarter.

For comparison units in sample (B) this procedure implies that they will only be matched to a treated unit if, in addition to being identical in the other covariates, they have an identical past 4-quarter labor force status history looking back from the point in time – the exact month – when the treated unit entered the program. Correspondingly, the 3-quarter outcome sequence for this matched comparison unit will be evaluated exactly congruent with the treated unit's 3-quarter post-treatment outcome sequence, i.e. after the treated person leaves the program. This approach accomplishes to define "treatment start" and "treatment stop" for comparison units, points in time that otherwise are not defined. Moreover, treated and untreated units are always compared during the same period, such that changes in general economic conditions, even on the local labor market level, are controlled for.[14]

Figure 1 illustrates the procedure for sample (A), which disregards the precise sequence of labor force states but imposes the timing requirements discussed above. The figure also shows how the 12-month-sequence is condensed into the comparable 4-quarter-structure. This method generates a "moving window" as the algorithm advances through the spells of treated units one after the other searching for comparable untreated units at the corresponding points in time. **Figure 2** demonstrates the increased stringency of the matching requirements associated with samples (B), where one or more controls are matched to a treated unit on the basis of identical pre-treatment labor force status sequences at the same point in time.

4.3 Pre-Treatment Histories

A central aspect of program evaluation regards the process that determines participation and non-participation in the program, and the potential problem of participants self-selecting into the treatment on the basis of observed or unobserved information. In the US, where ALMP measures were first evaluated, interest was mainly in the earnings

---

[14] Such changes did indeed occur in Poland during transition. For instance, overall unemployment displayed an inverted U-shape over our sampling period, increasing from 13.3% (1992) via 14.0% (1993) to 14.4% (1994), and then falling again to 13.3% (1995) and 12.4% (1996).

performance of participants. With respect to the determinants of participation, the focus then, logically, was on the differences in pre-treatment earnings performance of program participants and non-participants. In the context of controlling for such differences in the earnings performance of treatment and comparison groups prior to treatment, Ashenfelter (1978) pointed to a potentially serious limitation of difference-in-differences methods when he observed a relative decline in pre-treatment earnings for participants in subsidized training programs. This empirical regularity has been called "Ashenfelter's dip" and has been confirmed by subsequent analyses of many other training and adult education programs (cf. Bassi 1983, Ashenfelter and Card 1985, LaLonde 1986, Heckman, LaLonde, and Smith 1999).

For instance, Ashenfelter and Card (1985) apply a model that focuses on earnings changes as the determinants of participation. This line of thought was a logical consequence of both Ashenfelter's discovery and the main objective of the program, and resulted in analyses using earnings histories to eliminate differences between participants and non-participants. Clearly, the fact whether the pre-program earnings dip is transitory or permanent determines what would have happened to participants had they not participated, and the validity of any estimation approach depends on the relationship between earnings in the post-program period and the determinants of program participation (Heckman and Smith 1999).

This rather established observation that it is earnings dynamics that drive program participation has lately been put into question by Heckman and Smith (1999), who argue that it is rather labor force dynamics that determine participation in an ALMP program, a point they reinforce in their recent in-depth analysis of the determinants of program participation (Heckman and Smith 2004). This point had been made before by Card and Sullivan (1988), albeit implicitly, who analyze training effects conditional on pre-program employment histories. Furthermore, Heckman and Smith (1999) argue for a distinction between employment dynamics – indicating whether an individual is employed or not – and labor force dynamics, incorporating also whether a non-employed person is either unemployed or out-of-the-labor-force. Their conclusion is "that labor force dynamics, rather than earnings or employment dynamics, drive the participation process" (Heckman and Smith 1999). Therefore, we extend the "employment history setting" considered in Card and Sullivan (1988) to a "labor force status history setting", reflecting also movements in and out of inactivity. This approach is delineated above in section 4.2.

Figures 3 and 4 draw the distributions of pre-treatment labor market histories for samples (A) for both Training (Fig.3) and Intervention Works (Fig.4)[15]. Representing a 12-month labor force status sequence with 4 quarterly realizations of a trinomial variable (0,1,2) yields 81 possible sequences ("0000" to "2222"). For the purpose of illustrating the distributions – and only for that purpose – we classify these 81 sequences into 11 categories, so that on the abscissa the bottom categories contain "inactive" sequences (mostly "0"s), the middle categories comprise "unemployed" sequences ("2"s), and the top categories represent "employed" sequences ("1"s). Categories 1, 6, and 11 exclusively embody the straight sequences (i.e. "0000", "2222", and "1111", respectively).

Thus, of the three peaks we observe in the graphs in Figures 3 and 4, the left peak represents "inactive" histories, because histories with a low order number contain many "0"s. Accordingly, the peak in the middle expresses "unemployed" histories, and the peak to the right depicts "employed" histories. In terms of balancing of distributions, the picture is almost the same for Figures 3 and 4. For both policy measures samples (A) display only limited accordance in pre-treatment histories for treated and comparison units. The figures also show that treatment individuals in Training are quite different from those in Intervention Works. For the Training participants, the fractions of "employed" and "unemployed" histories are quite close to each other, while in the Intervention Works sample we observe a far larger fraction of "unemployed" histories among the treated. This is entirely in line with the regulations regarding selection for participating in Intervention Works (Section 2.2). Moreover, for both Training and Intervention Works the comparison samples (A) are too "successful" in that they contain too many "employed" sequences relative to "unemployed" sequences in order to be comparable to the treated units, where "unemployed" sequences dominate.

4.4 Dynamics of treatment assignment

Recent work on program evaluation has increasingly considered dynamic aspects inherent to estimating treatment effects of active labor market policies. Against the backdrop of generally rather disappointing evidence on the effectiveness of training across countries (Heckman et al. 1999, Kluve 2006), one recent line of research, for instance, shows that positive treatment effects may only materialize in the long run, and that program effectiveness can show a considerable dynamic ranging from often severe short-term

---

[15] Clearly, in sample (B) these distributions will be balanced.

locking-in effects to long-term gains in employment prospects (e.g. Lechner et al. 2004). Other lines of research on program evaluation dynamics discuss e.g. the evaluation of program sequences (Lechner and Wiehler 2007) or the treatment effects that arise from continuous variation in the program length (Kluve, Schneider, Uhlendorff, Zhao 2007).

Fredriksson and Johansson (2003) discuss dynamic program evaluation focusing on the timing of treatment as an outcome of a stochastic process. In their framework, a set of identical individuals enters unemployment at time zero, and then each individual does or does not start treatment at some subsequent point in time over the sampling frame. In particular, Fredriksson and Johansson (2003) discuss that by defining a comparison group as the group of individuals who were never treated (hence disregarding individuals who receive treatment later) one implicitly conditions on future outcomes, and they show that this can bias estimates towards finding negative treatment effects. Steiger (2005) is an application of this framework.

As explained in the previous sections, our approach models the dynamics of treatment assignment in a different way. We include all individuals who are unemployed during at least one quarter over the sampling period, which is the minimum requirement to be eligible for program participation at all. Then individuals can enter treatment at any point in time, and they can be unemployed at any point in time. This procedure creates sufficient flexibility for matching treated and comparison units at varying time points (the "moving window"), which is crucial to take into account the changes in institutional set-up (cf. section 2) and economic context common to a transition economy, which indeed actually occur over our sampling period. In addition, individual-level dynamics are captured in the pre-treatment labor force status sequences that delineate the idiosyncratic path to program entry.

5. Empirical results
5.1 Distributions of outcomes
Figures 5 and 6 plot distributions for the post-treatment employment success for treated units and comparison units in samples (A) and (B). There are 27 possible labor market status sequences capturing employment performance in the three quarters succeeding treatment. Similar to our presentation of pre-treatment labor market histories, we classify these 27 possible sequences of 3 quarterly realizations of a trinomial variable into 9 categories for illustration purposes. Once more, bottom categories contain "inactive"

sequences (category 1="000"), middle categories include "unemployed" sequences (category 5="222"), and top categories comprise "employed" histories (category 9="111"). Accordingly, in the graphs the left peak depicts "inactive" sequences, the middle peak "unemployed" sequences, and the right peak represents "employed" histories.

For the Training samples shown in Figure 5 we find, similar to what we have seen for the pre-treatment sequences of these samples (Figure 3), the "employed" and "unemployed" peaks have more or less the same height also for the post-treatment sequence. But while for sample (A) the "employed" peak is higher for comparison units than for treated units, and the "unemployed" peak is higher for treated units than for comparison units, this relation switches for sample (B). In (B) treated units display on average a slightly more successful post-treatment labor market sequence than corresponding comparisons. This would be an indication of a slightly – possibly insignificant – positive treatment effect of Training.

Looking at the Intervention Works samples in Figure 6, we find that in both samples the "unemployed" sequences are clearly predominant for the treated units. At the same time, comparison units display rather successful labor market histories in sample (A). In sample (B) this picture changes considerably, and a larger fraction of comparison units also displays "unemployed" histories. However, the comparison group still fares visibly better than the program participants. Sample (B) therefore indicates that during the 9 months directly succeeding participation in Intervention Works the treated units seem to be on average marginally – possibly insignificantly – less successful in finding employment than the comparison units.

Taken together, Figures 5 and 6 display three important patterns. First, moving from (A) to (B) we do not observe much variation in the distributions for treated units. Thus, the fact that we lose some treated units while increasing matching requirements does not seem to play an important role. Second, without conditioning on pre-treatment labor market histories the comparison samples apparently contain too many "successful" individuals – a pattern which we already observed for pre-treatment labor force status sequences in Figures 3 and 4. For samples (A) this would probably result in too negative an estimate of treatment effects. Third, across comparison units and treated units we observe clearly more "successful" outcomes for Training than for Intervention Works. This, too, is not surprising, as we noticed a similar relation for pre-treatment labor market history distributions (Fig. 3 and 4).

5.2 Treatment effect estimation

For purposes of the formal exposition of our estimation approach we consider a single generic intervention. Since we explicitly require that treated units be matched with comparison units from the identical set of observed pre-treatment and post-treatment months, any reference to the time period can be omitted from the formal exposition. In addition to the terminology introduced in section 3, let N_1 denote the number of treated units, with indices $i \in I_1$, and N_0 the number of potential comparison units, with indices $i \in I_0$. Potential labor market outcomes in post-treatment quarter q ($q = 1, 2, 3$) are denoted by Y_{qi}^1, if individual i received treatment, and by Y_{qi}^0, if individual i did not receive treatment. Outcomes are defined as multinomials with three possible realizations ("0"=out-of-the-labor-force, "1"=employed, "2"=unemployed), extending the formulations of Card and Sullivan (1988) from a binomial to a trinomial setting, i.e. labor force status histories instead of employment histories, and considering quarterly rather than yearly sequences.

We can only observe one of the two potential outcomes Y_{qi}^1 and Y_{qi}^0 for a given individual. This actual outcome is denoted by Y_{qi}. The objective is to formally construct an estimator of the mean of the unobservable counterfactual outcome $E(Y_q^0 | D=1)$. Since following the quarterly sequence of labor market outcomes might be too detailed for a direct economic interpretation of results, we condense the available information further and summarize the post-intervention labor market success of each individual i by the individual's average employment rate over the three quarters following the intervention[16]. Using an indicator function $\mathbf{1}(.)$, these employment rate outcomes are $\overline{Y}_i \equiv \frac{1}{3} \sum_q \mathbf{1}(Y_{qi} = 1)$, and $\overline{Y}_i^1$ and $\overline{Y}_i^0$, respectively, for employment rates with and without treatment. *Observed* outcomes for individual i can then be written as

$$(3) \qquad \overline{Y}_i = D_i \overline{Y}_i^1 + (1 - D_i)\overline{Y}_i^0,$$

and the impact of the intervention on the labor force status of individual i is given by

$$(4) \qquad \Delta_i = \overline{Y}_i^1 - \overline{Y}_i^0.$$

The parameters of interest in our evaluation analysis are weighted population averages over these individual treatment effects, the *mean effect of treatment on the treated* for types of

---

[16] Kluve et al. (1999) consider extensions to multinomial labor market states and longer post-treatment horizons.

individuals characterized simultaneously by specific sets of characteristics X; and labor force status histories before treatment h,

$$(5) \qquad E(\Delta | X, h, D=1) \qquad = \qquad E(\overline{Y}^1 - \overline{Y}^0 | X, h, D=1) \quad .$$

The ultimate interest there typically lies in the average treatment effects over the joint support of X and h given $D=1$,

$$(6) \qquad M = \sum_s w_s E(\Delta | s, D=1),$$

with s indicating any possible combination of X and h, and w_s representing the corresponding relative frequency in the treatment sample. In what follows we will also consider appropriate subsets of this joint support, and counterfactual treatment populations, using different relative weights.

How does this matching approach identify the parameters of interest? In randomized experiments the counterfactual expected values under no intervention can simply be estimated for intervention recipients by the mean values of the outcome for randomized-out would-be recipients. As detailed in section 3, matching methods can recover the desired counterfactual for a nonexperimental comparison group if unconfoundedness holds: Within each matched set of individuals, one can estimate the treatment impact on individual i by the difference over sample means, and one can construct an estimate of the overall impact by forming a weighted average over these individual estimates.

Matching estimators thereby approximate the virtues of randomization mainly by balancing the distribution of observed attributes across treatment and comparison groups, both by ensuring a common region of support for individuals in the intervention sample and their matched comparisons and by re-weighting the distribution over the common region of support. The central identification assumption is that of mean independence of the labor market status $\overline{Y}_i^0$ and of the treatment indicator D_i, given individual observable characteristics. In our specific application these conditioning characteristics are the demographic and regional variables X_i and the pre-treatment history h_i, i.e.

$$(7) \qquad E(\overline{Y}^0 | X, h, D=1) \qquad = \qquad E(\overline{Y}^0 | X, h, D=0) \quad .$$

We believe that conditioning on both core socioeconomic characteristics – including, in particular, information on local labor markets – and detailed labor force status sequences with exact alignment of the pre-treatment period lends plausibility to the unconfoundedness assumption, reflecting a meticulous adjustment of the method to the

research question and data at hand. Moreover, using the longitudinal structure of the data, labor force status sequences likely reflect relevant unobserved but time-persistent differences, such as e.g. motivation, between treated and untreated individuals.

In a standard difference-in-differences approach pre-treatment and post-treatment outcomes are typically treated symmetrically and the identifying assumption is that the change in outcomes that treated individuals would have experienced had they not received treatment would have been the same change – on average – that untreated individuals experienced during the same period. Whereas this assumption accounts for the phenomenon that participants typically experience lower pre-treatment outcomes, even though they might otherwise be identical to comparison units, it does not lend itself naturally to the analysis of categorical outcome variables. In this context, a natural generalization of the difference-in-differences idea is to condition on the specific realization of the outcome variable in the pre-treatment period, as we do in our study. This is possible, since due to the categorical nature of the outcome the conditioning remains tractable.

Our matching estimator applies an oversampling exact covariate matching within calipers, allowing for matching-with-replacement. Our particular attention to pre-treatment labor force status sequences implements this idea of a generalized difference-in-differences juxtaposition between treated units and comparison units. Due to the relevance of the previous history for subsequent labor market success – state dependence is one of the issues most discussed in the labor literature – we also emphasize this variable in the construction of the estimates. Specifically, for any treatment group history h for which at least one match could be found, we estimate the impact of the intervention by

$$(8) \qquad \hat{M}_h \;=\; \frac{1}{N_{1h}} \sum_{i \in I_{1h}} \left[\overline{Y}_i^1 - \sum_{j \in I_{0h} \wedge X_j \in C(X_i)} \frac{1}{n_{i0}} \overline{Y}_j^0 \right] \;,$$

where N_{1h} is the number of individuals with history h who receive the intervention ($N_1 = \sum_h N_{1h}$), I_{1h} is the set of indices for these individuals, $C(X_i)$ defines the caliper for individual i's characteristics X_i, and n_{i0} is the number of comparisons with history h who are falling within this caliper, with the set of indices for comparison units with history h being I_{0h}. The estimated standard error of the treatment effect estimate is then constructed as a function of the underlying estimated probabilities. The overall effect of the intervention is estimated in a last step by calculating a weighted average over the history-specific

intervention effects,

$$(9) \qquad \hat{M} \quad = \quad \sum_h w_h \hat{M}_h \quad .$$

using the treated units' sample fractions $N_{1h} \big/ \sum_h N_{1h}$ as weights. One can easily employ alternative weights – for instance the treated units' sample fractions observed for another intervention. This will allow to contrast the performance of alternative measures holding the composition of the treatment population constant.

5.3 Treatment effect results

Table 5 presents average treatment effects on the post-intervention employment rate for Training. The treatment effect estimate for sample (A) is insignificant, while the estimate obtained from sample (B) indicates that participation in Training results in an employment rate that is on average nearly 14 percentage points higher than it would have been in the absence of the program. Looking more closely at a classification by labor market history, we find the "peaks" from Figure 3, indicating that the share of "1111" sequences is almost as large as the share of "2222" sequences. Subsample sizes here are too small to draw any firm conclusions. Yet, it becomes transparent that the imbalance in unobservables that a matching algorithm would have produced, had it not accounted for labor market histories, would have emphasized the relatively "successful" nature of comparisons. Thus, it would have given the program less credit for the employment success of participants than would have been appropriate.

Stratifying the sample by time of entry into training allows us to make the point that the difference in the employment rates of the treated and the controls is unlikely driven by benefit regulations. Those who entered training before January 1st, 1995 were entitled to a full round of benefits if they could not find a regular job after the intervention, while this generous provision was cancelled for training participants entering after this date. The treatment effect is, however, larger - and statistically significant – for the subsample of entrants of the earlier period.

To illustrate the performance of the algorithm, we also report two "raw effects" resulting from simpler matching variants. The first one reports the effect we would estimate on a sample only using the timing structure, i.e. the moving window, but no information on covariates and labor force status sequences. The second one reports the

effect obtained from a simple covariate matching, without use of the moving window.[17] The results for training show that both of these effects are similar to the one obtained from Sample (A) (and all are insignificant), and that including the labor force status histories makes the difference in revealing the effectiveness of the training scheme.

Table 6 reports the treatment effect estimates for Intervention Works. It is to be expected that, on the basis of sample (A), estimates are more negative than the estimate derived from sample (B). After all, the relatively favorable composition of the comparison group with respect to unobservables is reflected in the pre-treatment labor market histories documented in Figure 4. Classification by labor market history allows us to look at the two major labor force status sequences that drive the peaks observed in Figure 4. For "employed" (1111) histories subsample sizes are rather small and the effects not well defined. For the subsample of "unemployed" (2222) histories, which entails almost 80% of total treated and comparison units, we find a significantly negative treatment effect close to the full sample effect. This is certainly no surprise, as the estimate of the full sample effect is dominated by the "2222" subsample effect. The "raw effects" reported in Table 6 illustrate the importance of including the moving window structure, but above all they show again how crucial it is to control for individual labor force status histories.

In finding reasons for the negative treatment effects of Intervention Works, it is sometimes suggested that subsidized jobs are of lower quality, locking the participating workers in a dead end, rather than preparing them for future labor market success. It might also be a stigmatization effect that causes participants of an employment program like Intervention Works to perform worse in the labor market than non-participants. Prospective employers might identify participants as "low productivity workers" and are not willing to accept them into regular jobs. Another explanation, which might have particular merit in the Polish case, is "benefit churning": Workers with long unemployment spells who have difficulty finding regular employment might be identified by labor bureau officials and only be chosen for participation in the Intervention Works scheme so that they re-qualify for another round of benefit payment. It is perhaps not a mere coincidence that the large majority of Intervention Works jobs in our sample lasts six months, exactly the length of time one needs to work within the year preceding benefit receipt in order to

---

[17] The number of untreated observations used in estimating this second raw effect is smaller than the one used for the first raw effect due to restrictions necessary for calculating the respective effects. In particular, since no flexible timing structure is imposed for the second raw effect, the average employment rates during the last nine months of the sampling period are used to calculate the effect. Because for the first raw effect this is flexible over the sampling period and need not be the last nine months, more observations can be used.

qualify for unemployment benefits. This is illustrated by the average program duration of 6.1 months (Table 4) – in fact, more than 57% of participants stayed in Intervention Works for exactly six months, and about 79% for six months or longer.

If selection into Intervention Works would indeed depend on the outcome (i.e. becoming unemployed again and receiving benefits after the program ends) in this manner, it might be problematic to maintain the unconfoundedness assumption. To address this aspect, however, we include an additional covariate indicating whether the individual received benefits in the last month before entering the program. This indicator – especially in conjunction with the labor force status history – captures the dynamic of running out of benefits (while remaining unemployed) before program start, and thus controls for benefit exhaustion being a potential selection criterion.

The corresponding results shown at the bottom of Table 6 are similar to the ones obtained without the benefit indicator and continue to point to a generally negative effect of Intervention Works. For Sample (A) the estimated treatment effect is less negative than without the indicator, which likely reflects the fact that the indicator already captures some aspect of the labor market history. The estimate based on Sample (B) is very close to the one without the benefit receipt indicator (a decrease in the employment probability of 12.6 and 14.7 percentage points, respectively), though the number of matched treated units is reduced quite a bit.

In addition to the effects of a program *per se*, i.e. relative to a no-program state, it is often of interest for policy makers to compare programs, i.e. to learn about how effective one particular program is relative to some other program. This issue has been addressed by the literature on "multiple treatments", which extends the Potential Outcome Model to settings in which treatment can take on more than two values (cf. Imbens 2000, Lechner 2001 and 2002 for discussion). This extends the set of relevant counterfactuals, in particular assessing the question "What would have happened to person *i*, who participated in program A, had she instead participated in program B?".

In the spirit of this literature, **Table 7** presents treatment effect estimates for samples (B) obtained from a simple "counterfactual experiment". The first line reports the factual Training treatment effect estimate computed as shown in Table 5. This estimate tries to answer the question: "How much did Training participants benefit from participating in Training?" The second line reports a "counterfactual" Intervention Works treatment effect for Training participants, i.e. it tries to answer the question: "How much

would Training participants have benefited, if they had instead participated in Intervention Works?" The estimate is obtained by history-wise reweighting the Intervention Works sample using the fraction of the treated units in the Training sample as weights. Clearly, this reweighting by labor market history implicitly assumes that there are no relevant changes in other elements of X, and emphasizes that the labor force status histories are the crucial characteristic determining program effectiveness. As Table 4 has shown, Training and Intervention Works participants are relatively similar in core covariates, with the exception of education. The main difference between the two program groups are in the distribution of the labor force status sequences (recall Figures 3 and 4). Hence, in the context of our analysis focusing on the importance of these sequences, we are interested in relative program effects that arise from counterfactual distribution of labor force histories.

The estimate in the second row of Table 7 shows that, while the Intervention Works effect on Training participants still displays a negative sign, the effect is insignificant, so that Training participants participating in Intervention Works would have done better than Intervention Works participants themselves. Looking at the effects of Training and Intervention Works, respectively, on Intervention Works participants, we find the counterpart to this result: Intervention Works participants participating in Training instead would not have gained as much from the treatment as Training participants themselves. Thus, persons with better observable and unobservable characteristics seem to have been targeted for the Training program.

The last two lines in Table 7 report differential treatment effects of Intervention Works vs. Training. The estimates represent the difference between the difference of treated and comparison units in Intervention Works and the difference of treated and comparison units in Training. Once more, differences are taken history-wise and weighted using either Intervention Works participants or Training participants sample weights. Both estimates clearly show that Training is the superior ALMP relative to Intervention Works.

The methodology used in our paper allows us to evaluate ALMP at the individual level. It tells us that those persons participating in Polish Training programs have better employment prospects than they would have had had they not participated and also that they have better employment prospects than those who take part in Intervention Works. The methodology does not address the issue whether Training improves the overall performance of the labor market, i.e., for example, whether it lowers the aggregate unemployment rate. Even if Training is beneficial at the individual level, substitution

effects – Training participants just "jump the queue" of those in line for regular jobs – could neutralize its impact at the aggregate level. On the other hand, the finding that a program is not even effective at the individual level, like the Polish Intervention Works scheme, helps us to focus attention on targeting issues and/or wrong incentive structures that distort the behavior of labor bureau officials and of the unemployed.

6. Conclusion

Over the last decade, there has been much interest by labor economists and econometricians – from both practical and theoretical perspectives – in the evaluation of treatments, in particular labor market interventions such as training and wage subsidy schemes. Building on a rather established culture of implementing and evaluating such Active Labor Market Policies in the US, Western European countries have caught up substantially in this regard. The next to follow are countries of Eastern Europe, often called "transition countries", that in the early years of transition in many cases implemented unemployment benefit schemes and active labor market programs similar to Western welfare systems. In the case of Poland, even though programs had to be built from scratch, already in the year 1996 data were collected that allow comprehensive evaluation of the country's ALMP measures during the first half of the 1990s.

In this study, we present an evaluation of two Polish active labor policies, Training and Intervention Works, i.e. a wage subsidy scheme. There is a set of features that we think make this study especially interesting. First, we use data from the 18th wave of the Polish Labor Force Survey that contain extensive and detailed information on individual labor force status histories from January 1992 until August 1996. Second, the evaluation is set against the background of a country in the early years of transition, a fact that implies a rapidly changing macroeconomic environment, making it indispensable to develop a treatment effect estimator that can account for these changes appropriately. Third, in addition to estimating treatment effects on the basis of individual labor force status sequences, we use the detailed monthly data to investigate further how important such labor market histories are in fact for determining participation in the program.

The core part of the analysis is the development of a matching estimator based on individual pre-treatment labor force status sequences. This creates a "moving window" structure that allows for individually flexible entry into and exit out of the program, hence conditioning on covariates and labor force status histories at exactly the month of program

start, and comparing outcomes at exactly the month of program termination. Clearly, while increasing comparability of treated and comparison units, this procedure also nicely controls for changes in the macroeconomic environment. Like in many transition countries, we are confronted in the Polish case with a rapidly changing environment of labor market institutions and regulations. Using a "moving window" structure is, therefore, also vital in ensuring that the estimated treatment effects are not driven by benefit regulations or labor market institutions affecting the treated and the control units in a different fashion.

Our approach is delineated using two matched samples (A) and (B) that differ in matching requirements. In terms of treatment effect estimates we find a positive effect of training on the treated population, and a negative effect of Intervention Works. Our results on training add to an increasing number of evaluation studies across European countries, including some accession countries, that identify training programs as one Active Labor Market Policy that can be effective.[18] The evidence on wage subsidy schemes is mixed, ranging from often very successful programs (e.g. in Germany and Switzerland) to sometimes less promising results, as is also the case here. At any rate, we feel that our results on Intervention Works is specific to the Polish case, as characterized by a particular incentive structure, including e.g. the possibility of a cycling behavior to renew benefits.

Most importantly for our paper, we document that pre-treatment labor force status sequences contain indispensable information regarding selection into treatment and that controlling for these histories can eliminate a large part of the overt bias between treated and comparison units. This result confirms and reinforces the point made by Heckman and Smith (1999, 2004) in a completely different labor market.

---

[18] See Kluve (2006) for an assessment of training measures in Western European countries; studies demonstrating the efficacy of training measures in accession countries are, e.g., Kotzeva (2000) for Bulgaria, Lubyova and van Ours (1999) for Slovakia, and Leetma and Võrk (2004) for Estonia.

References

Ashenfelter O. Estimating the Effect of Training Programs on Earnings. Review of Economics and Statistics 1978; 60; 47-57.

Ashenfelter O, Card D. Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. Review of Economics and Statistics 1985; 67; 648-660.

Bassi L.J. The Effect of CETA on the Post-Program Earnings of Participants. The Journal of Human Resources 1983; 18; 539-556.

Card D ,Sullivan, D. Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment. Econometrica 1988; 56; 497-530.

Dehejia R, Wahba S. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. Journal of the American Statistical Association 1999; 94; 1053-1062.

Dehejia R, Wahba S. Propensity Score-Matching Methods for Nonexperimental Causal Studies. The Review of Economics and Statistics 2002; 84; 151-161.

Dehejia R. Practical Propensity Score Matching: a Reply to Smith and Todd. Journal of Econometrics 2005; 125; 355-364.

Fisher R.A. The Design of Experiments. Oliver & Boyd: Edinburgh; 1935.

Fredriksson, P, Johansson P (2003), Program Evaluation and Random Program Starts, IFAU Working Paper 2003:1, Uppsala.

Golinowska, S. Ochrona socjalna bezrobotnych w Polsce oraz w innych krajach (Social Protection of the Unemployed in Poland and in other Countries). CASE Studies and Analyses No. 201, Warsaw; 1999.

Góra M, Schmidt CM. Long-term Unemployment, Unemployment Benefits and Social Assistance: The Polish Experience. Empirical Economics 1998; 23; 55-85.

Heckman JJ, Ichimura H, Todd PE. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. Review of Economic Studies 1997; 64; 605-654.

Heckman, JJ. LaLonde, RJ, Smith, JA 1999. The Economics and Econometrics of Active Labor Market Programs. In: Ashenfelter, O, Card, D (Eds). Handbook of Labor Economics, vol. III. North-Holland: Amsterdam; 1999.

Heckman JJ, Smith JA. The Pre-programme Earnings Dip and the Determinants of Participation in a Social Programme: Implications for Simple Programme Evaluation Strategies. The Economic Journal 1999; 109; 313-348.

Heckman JJ, Smith JA. The Determinants of Participation in a Social Program: Evidence from a Prototypical Job Training Program. Journal of Labor Economics 2004; 22; 243-298.

Holland PW. Statistics and Causal Inference (with discussion). Journal of the American Statistical Association 1986; 81; 945-970.

Imbens, GW. The Role of the Propensity Score in Estimating Dose-Response Functions, Biometrika 2000; 87; 706-710.

Imbens GW. Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. Review of Economic and Statistics 2004; 86; 4-29.

Kluve J. The Effectiveness of European Active Labor Market Policy. RWI: Discussion Papers 2006; 37.

Kluve J, Lehmann H, Schmidt CM. Active Labor Market Policies in Poland: Human Capital Enhancement, Stigmatization, or Benefit Churning? Journal of Comparative Economics 1999; 27; 61-89.

Kluve, J, Schmidt, CM. Can Training and Employment Subsidies Combat European Unemployment? Economic Policy 2002; 35; 409-448.

Kluve, J, Schneider H, Uhlendorff A, Zhao Z. Evaluating continuous training programs using the generalized propensity score, November 2007, unprocessed.

Kotzeva, M. Evaluation of Active Labor Market Programs in Bulgaria, 2000, unprocessed.

Kubiak, P. Efekty uczestnictwa bezrobotnych w aktywnych programach rynku pracy w Polsce (The Effects of Participation of the Unemployed in Active Labor Market Policies in Poland). University of Łódz Press: Łódz; 2005.

LaLonde RJ. Evaluating the Econometric Evaluations of Training Programs With Experimental Data. American Economic Review 1986; 76; 604-620.

Lechner, M. Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption. In: M. Lechner and F. Pfeiffer, eds., Econometric Evaluation of Labour Market Policies. Physica: Heidelberg; 2001.

Lechner, M. Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies. Review of Economics and Statistics 2002; 84; 205-220.

Lechner, M, Miquel, R, Wunsch, C. Long-Run Effects of Public Sector Sponsored Training in West Germany. IZA Discussion Paper 1443, 2004.

Lechner, M, Wiehler SA. Does the Order or Timing of Active Labour Market Programs matter? University of St. Gallen 2007, unprocessed.

Leetma, R., Võrk, A. Evaluation of Active Labour Market Programmes in Estonia, 2004, unprocessed.

Lubyova, M., vanOurs, J. Effects of Active Labor Market Programs on the Transition Rate from Unemployment into Regular Jobs in the Slovak Republic. Journal of Comparative Economics 1999; 27; 90-112.

Neyman J. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9., translated and edited by D.M. Sabrowska and T.P. Speed from the Polish original, which appeared in Roczniki Nauk Rolniczych Tom X (1923), 1-51 (Annals of Agriculture), Statistical Science 1990; 5; 465-472.

OECD. Social Expenditure Database 1980-1997, 2nd ed. OECD: Paris; 2000.

OECD. Employment Outlook. OECD: Paris; 2004.

Puhani P. Advantage through Training? A Microeconometric Evaluation of the Employment Effects of Active Labour Market Programmes in Poland. ZEW Disc. Paper 1998; 98-25.

Rosenbaum PR. Observational Studies. Springer Series in Statistics: New York; 1995.

Rubin DB. Estimating Causal Effects of Treatments in Randomized and Nonrandomized

Studies. Journal of Educational Psychology 1974; 66; 688-701.

Rubin DB. Assignment to Treatment Group on the Basis of a Covariate. Journal of Educational Statistics 1977; 2; 1-26.

Smith JA, Todd, PE. Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? Journal of Econometrics 2005a; 125; 305-353.

Smith JA, Todd, PE. Rejoinder. Journal of Econometrics 2005b; 125; 365-375.

Steiger, H. Is less more? A look at nonparticipation in Swiss active labour market programmes. University of St. Gallen 2005, unprocessed.

Table 1. Expenditures of the Labor Fund (LF) and on active measures

Year	Expenditures of LF		Expenditures on active measures	
	mln zł	% of GDP	mln zł	% LF
1990	370.2	0.70	118.7	32.1
1991	1358.4	1.80	94.7	7.0
1992	2282.7	2.15	107.5	4.7
1993	3190.3	2.31	354.5	11.1
1994	4447.1	2.46	569.0	12.8
1995	6328.3	2.55	754.9	11.9
1996	7525.3	2.41	806.1	10.7
mean		2.05		12.9

Source: Kubiak (2005) and own calculations.

Table 2. Expenditures on Training and Intervention Works

Year	Training		Intervention works	
	mln zł	percent*	mln zł	percent*
1990	1.6	1.3	20.9	17.6
1991	9.0	9.5	45.3	47.8
1992	19.3	18.0	46.9	43.6
1993	45.0	12.6	136.6	38.3
1994	59.5	10.5	231.7	40.7
1995	63.1	8.4	311.1	41.2
1996	86.1	10.7	277.2	34.4

* Percent of total expenditures on active measures. Source: Kubiak (2005).

Table 3. Number of participants in all measures and in training and intervention works 1992-1996*

Participants in	1992	1993	1994	1995	1996
All measures in thousands	218.1	295.7	412	437.2	367.3
% of unemployed	8.7	10.2	14.5	16.6	15.6
Training in thousands	70.2	75.8	91.7	84.1	86
% of unemployed	2.8	2.6	3.2	3.2	3.7
Intervention Works	104.5	132.4	195.4	206.3	139.6
% of unemployed	4.2	4.6	6.9	7.8	5.9

* End-of-year stocks. Source: Own calculations based on data from Ministry of Labor and Social Policy and Central Statistical Office (GUS).

Table 4. Composition of matched samples

		Training	Intervention Works
Reservoir	age	34.5	36.3
	%educated[a]	91.7	64.0
	%female	56.2	40.4
	%married	66.9	67.6
	# treated	121	275
	# untreated	6751	6757
	Avg. program duration for treated	2.51	6.11
Sample (A)	# treated	114	244
	# untreated	983	1354
Sample (B)	# treated	87	212
	# untreated	111	240
	[# untreated][b]	110	211

[a] Excluding individuals with only primary school attainment or less.
[b] Number of observations that the algorithm matched exactly once.

Table 5. Average post-treatment employment rates – Training

	treated units	comparison units	effect	std.error
Sample (A)	114	983	-.048	.049
Sample (B)	87	111	.138	.059
Sample (B) stratified by labor force status history:				
"1111"	24	34	.071	.115
"2222"	32	43	-.077	.103
Sample (B) stratified by program entry date				
Before Jan 1, 1995	55	73	.152	.078
After Jan 1, 1995	32	38	.122	.110
Raw effect (i): No covariates	121	6751	-.027	.046
Raw effect (ii): No moving window	121	6309	-.040	.045

Table 6. Average post-treatment employment rates – Intervention Works

	treated units	comparison units	effect	std.error
Sample (A)	244	1354	-.291	.031
Sample (B)	212	240	-.126	.040
Sample (B) stratified by labor force status history:				
"1111"	16	19	.084	.148
"2222"	168	191	-.150	.045
Raw effect (i): No covariates	275	6757	-.285	.026
Raw effect (ii): No moving window	275	6322	-.312	.030
Additional covariate: benefit receipt				
Sample (A)	242	1152	-.208	.033
Sample (B)	149	243	-.147	.037

Table 7. Counterfactual treatment effects

Treatment	Weights	Effect	Std.Error	Interpretation
Training	Training	.138	.059	Factual Training treatment effect
Intervention Works	Training	-.048	.064	Counterfactual Intervention Works treatment effect
Intervention Works	Intervention Works	-.126	.040	Factual Intervention Works treatment effect
Training	Intervention Works	.089	.083	Counterfactual Training treatment effect
Intervention Works – Training	Intervention Works	-.218	.093	Differential treatment effect Intervention Works vs. Training
Training – Intervention Works	Training	.185	.087	Differential treatment effect Training vs. Intervention Works

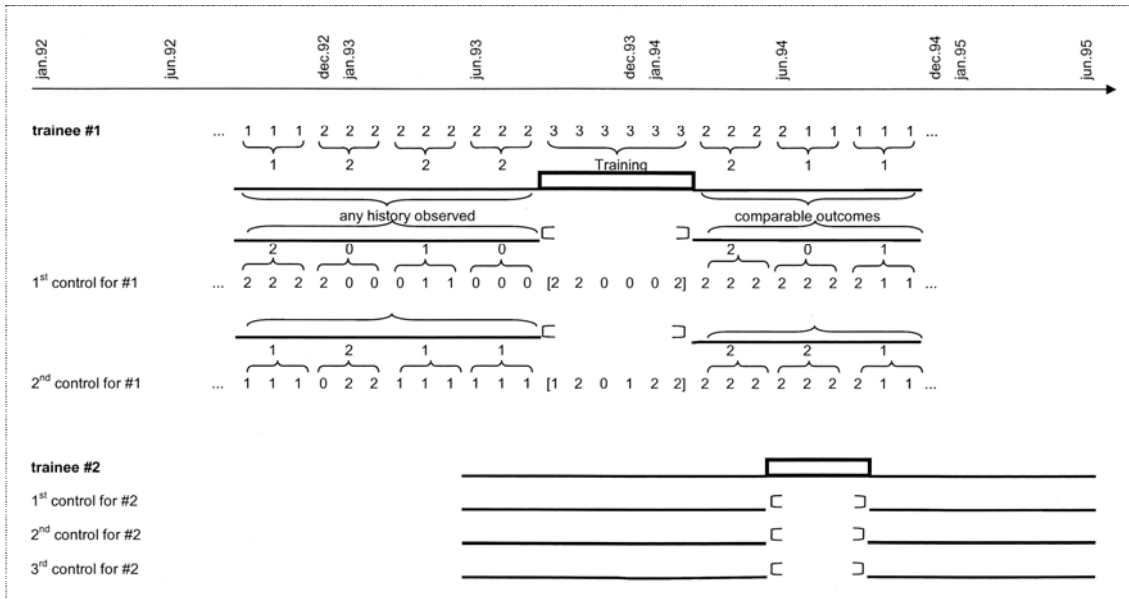**Figure 1. Matching using a "moving window" in Sample (A)**

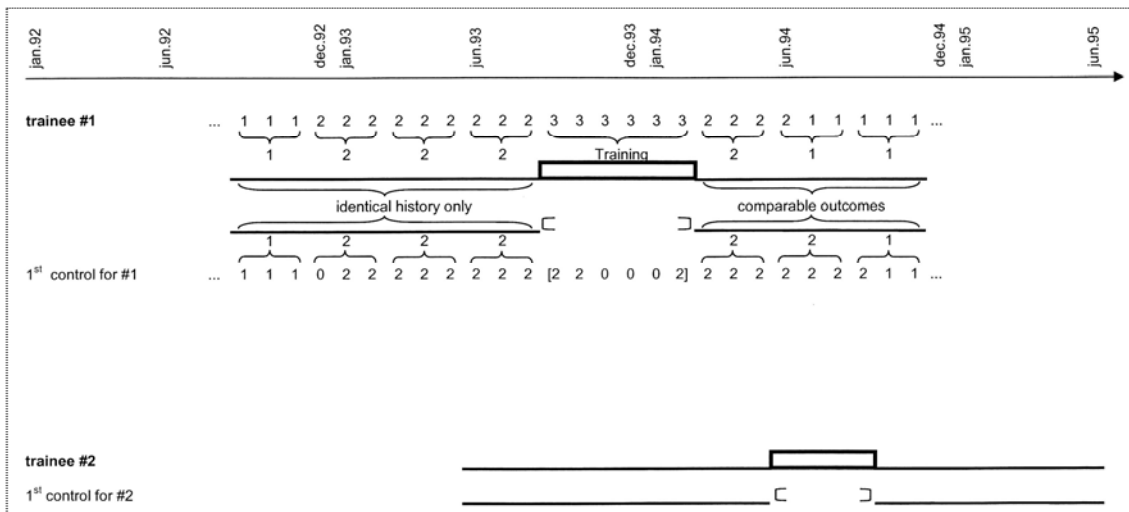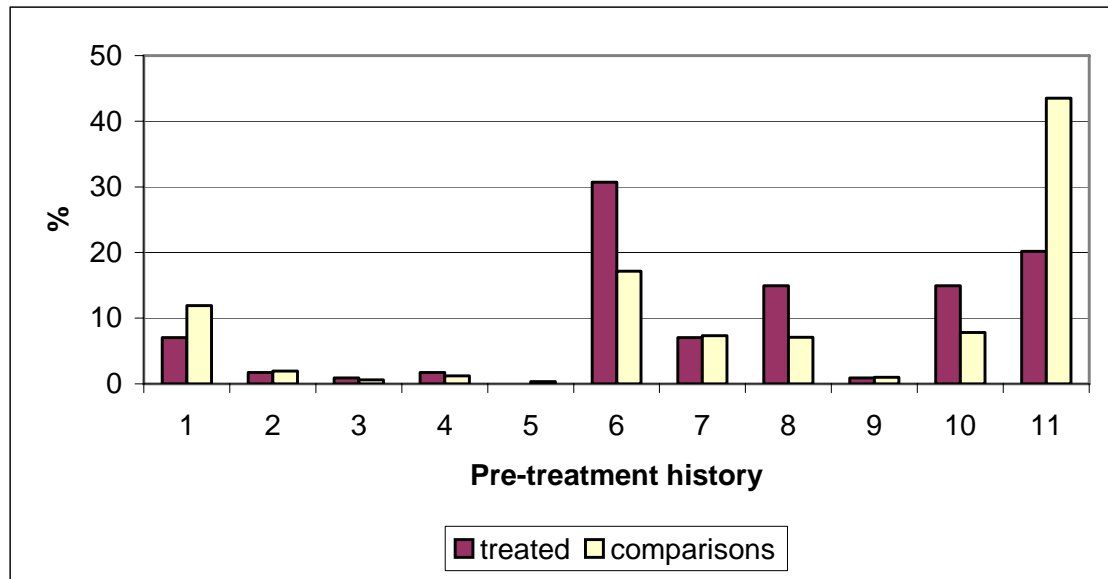**Figure 2. Matching using labor market histories and a "moving window" in Sample (B)**

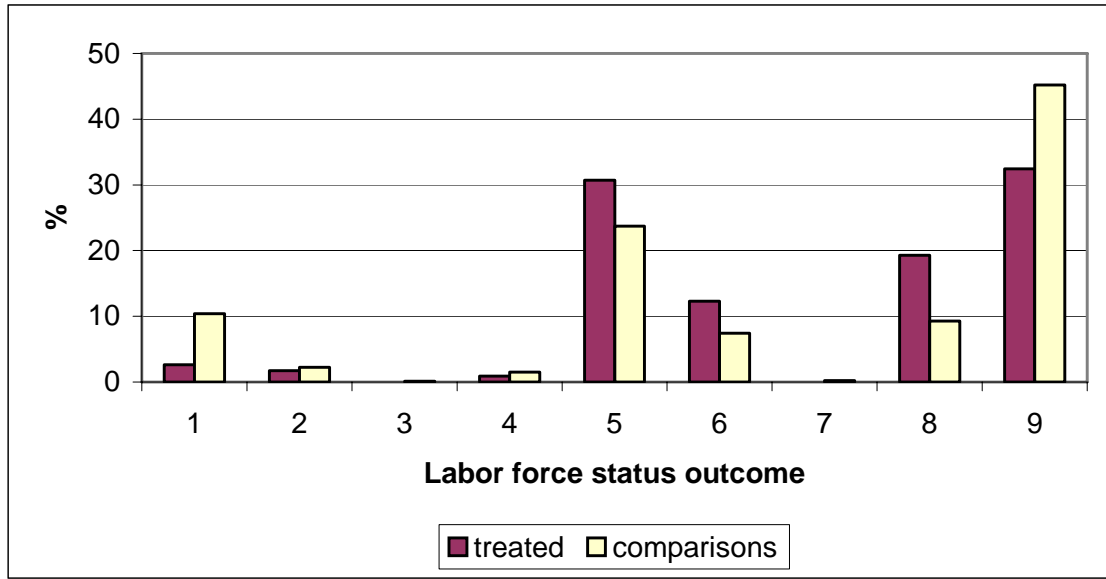**Figure 3. Distribution of pre-treatment labor market histories – Training sample (A)**

The 3⁴ possible labor force status sequences are classified into 11 categories: "1"=0000; "2"=0001, 0010, 0100, 1000, 0002, 0020, 0200, 2000; "3"=0012, 0102, 1002, 0120, 1020, 0021, 1200, 0201, 0210, 2100, 2010, 2001, 0110, 1010, 1100;"4"=0022, 0202, 2002, 0220, 2020, 2200; "5"=2201, 2021, 0221, 2210, 2012, 0212, 2120, 2102, 0122, 1220, 1202, 1022; "6"=2222; "7"=2220, 2202, 2022, 0222, 2221, 2212, 2122, 1222; "8"=2211, 2121, 1221, 2112, 1212, 1122; "9"=1102, 1012, 0112, 1120, 1021, 0121, 1210, 1201, 0211, 2110, 2101, 2011, 0011, 0101, 1001; "10"=1110, 1101, 1011, 0111, 1112, 1121, 1211, 2111; "11"=1111.

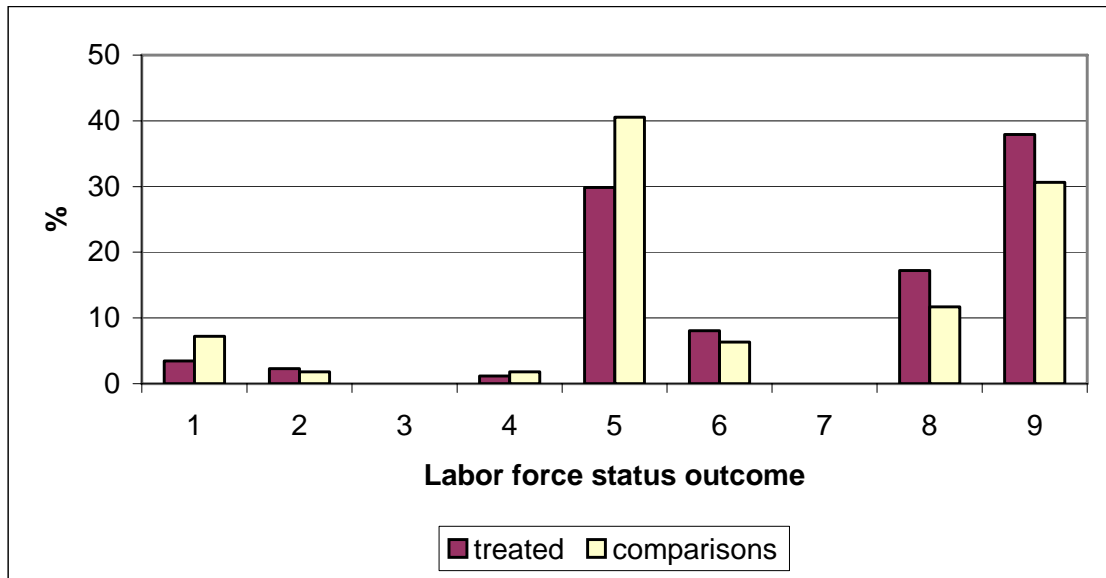**Figure 4. Distribution of pre-treatment labor market histories – Intervention Works sample (A)**
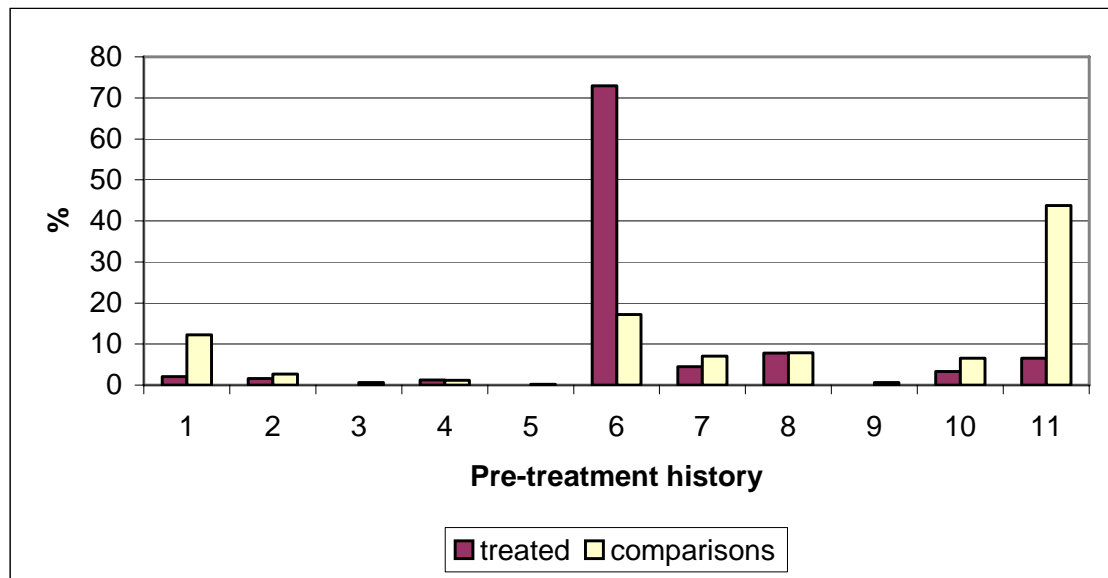
The 3⁴ possible labor force status sequences are classified into 11 categories: "1"=0000; "2"=0001, 0010, 0100, 1000, 0002, 0020, 0200, 2000; "3"=0012, 0102, 1002, 0120, 1020, 0021, 1200, 0201, 0210, 2100, 2010, 2001, 0110, 1010, 1100;"4"=0022, 0202, 2002, 0220, 2020, 2200; "5"=2201, 2021, 0221, 2210, 2012, 0212, 2120, 2102, 0122, 1220, 1202, 1022; "6"=2222; "7"=2220, 2202, 2022, 0222, 2221, 2212, 2122, 1222; "8"=2211, 2121, 1221, 2112, 1212, 1122; "9"=1102, 1012, 0112, 1120, 1021, 0121, 1210, 1201, 0211, 2110, 2101, 2011, 0011, 0101, 1001; "10"=1110, 1101, 1011, 0111, 1112, 1121, 1211, 2111; "11"=1111.

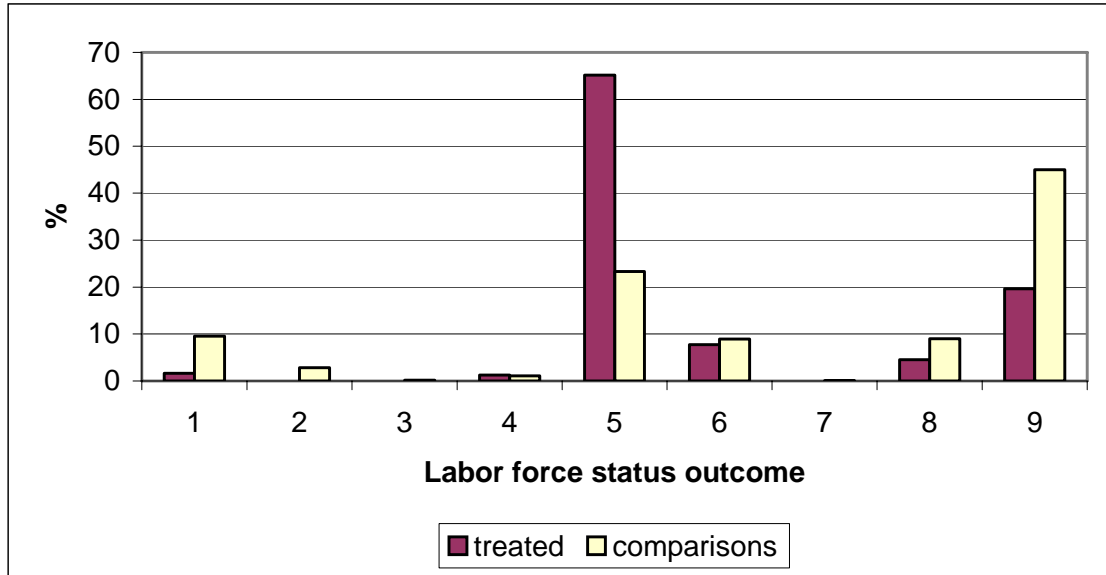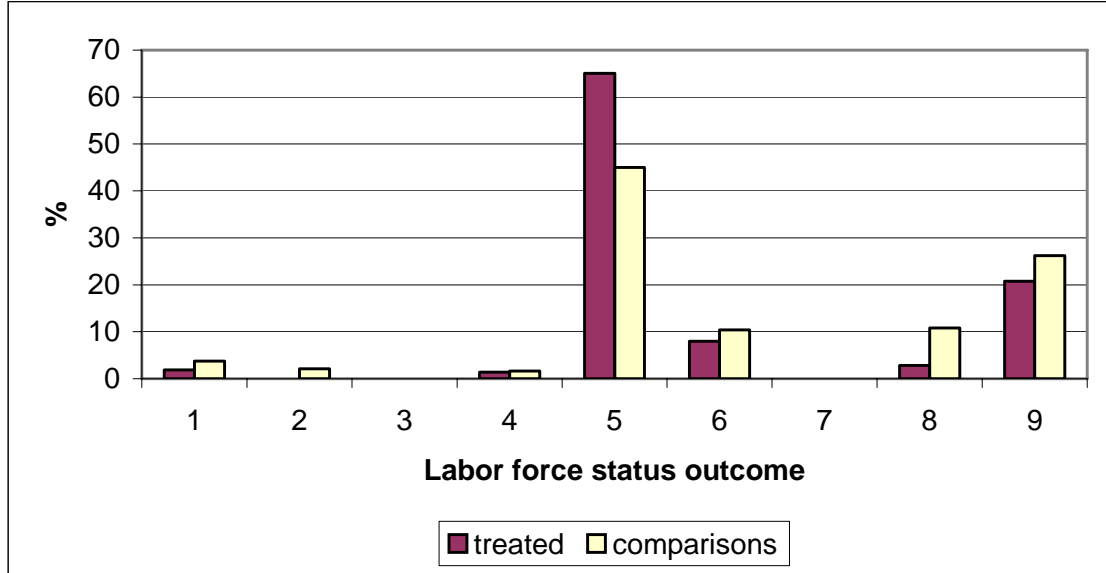**Figure 5. Distribution of post-treatment labor market sequences – Training Sample (A)**

Sample (B)

The 3³ possible labor force status sequences are classified into 9 categories: "1"=000; "2"=001, 010, 100, 002, 020, 200; "3"=210, 120, 102; "4"=220, 202, 022; "5"=222; "6"=221, 212, 122; "7"=012, 021, 201; "8"=110, 101, 011, 112, 121, 211; "9"=111.

Figure 6. Distribution of post-treatment labor market sequences – Intervention Works
Sample (A)

Sample (B)

The 3³ possible labor force status sequences are classified into 9 categories: "1"=000; "2"=001, 010, 100, 002, 020, 200; "3"=210, 120, 102; "4"=220, 202, 022; "5"=222; "6"=221, 212, 122; "7"=012, 021, 201; "8"=110, 101, 011, 112, 121, 211; "9"=111.