

Christopher J. O’Leary

Senior Economist
W.E. Upjohn Institute for Employment Research
Kalamazoo, Michigan, USA

EVALUATING ACTIVE LABOR MARKET PROGRAMS IN TRANSITION ECONOMIES

1. Introduction¹

The theme of this conference, “Active Labor Market Programs: Improvement of Effectiveness“ is a topic of central importance to human resource policy makers. The conference aim is to help shape an evaluation strategy for government employment programs in the Russian Federation. Over the past decade, several economies in transition from central planning to market based resource allocation systems have devoted significant effort to measuring the effectiveness of employment programs.

This paper provides an overview of the important issues involved in evaluation studies of employment programs. Among economies in transition, my longest personal experience has been in Hungary, where the shift to a market economy is at an advanced stage. Examples of evaluation experience in Hungary are used in this paper to illustrate the alternative methods of evaluating active labor market programs.

The next section of the paper provides a simple statement of the alternative types of outcome concepts which may be examined. These ideas are the foundation for the discussion of performance monitoring and net impact estimation provided in the next two sections. Explanations of each approach to evaluation includes enumeration of ways to use the results along with caveats on interpretation of findings. The conclusion briefly considers the macro-economic and political context of using evaluation results.

My conference remarks will supplement this prepared text in a few ways. I will mention special issues associated with doing evaluations in the transition context and alternative approaches. I will also suggest a unified strategy for developing an evaluation system for active labor market programs which will support management, planning and program development.

2. Concepts in Evaluation

In considering evaluations of employment programs it is important to be clear about the distinct concepts which may be examined. In terms of program outcomes three main types are of interest: gross outcomes, gross impacts, and net impacts. A gross outcome is simply the mean of an outcome of interest among program participants. A gross impact is the difference between program participants and non-participants on an outcome of interest. Gross impacts are of little use in understanding program effectiveness, and can be misleading to program management and policy decisions. Net impacts are the difference between mean outcomes of a representative sample of program participants and an appropriate sample of persons not receiving services. Great care must be taken in forming the

¹This paper draws on personal experience evaluating active labor market programs with World Bank funding in Hungary, Poland, and China. It is also influenced by methods used in my work for the U.S. Department of Labor, Human Resources Development Canada, and the International Labor Office. The latter work includes the recent publication by O’Leary, Nesporova and Samorodov (2001). The outline of the paper and conference remarks closely follow the structure of an address I presented to the World Association of Public Employment Services (WAPES) conference in Budapest, Hungary, during March, 2000.

latter group, which is called the comparison group. Proper net impact estimation can be done through random assignment in experimental studies, or by using statistical means to mimic the ideal of an experiment.

To firmly set distinct outcome concepts, consider a program intended to improve the chances of reemployment. Among program participants and the comparison group we may examine the rate of reemployment. Suppose that the rate of reemployment among program participants is 60 percent, that the observed rate among all previously unemployed is 40 percent and, that the rate among an appropriately chosen comparison group is 50 percent. In this example the program gross outcome is 60 percent, the program gross impact is 20 percent, and the program net impact is 10 percent.

The two most popular evaluation techniques for employment programs are *performance monitoring*--usually of gross outcomes, and *net impact estimation*. Net impact estimation is ideally conducted through *classically designed field experiments*. However, usually net impact estimation done by a cheaper and quicker *quasi-experimental method* which relies on statistical methods to mimic an experiment.

3. Performance Monitoring

Performance monitoring of gross program outcomes is usually done as part of a management system with an annual cycle. The process to develop and use such a system should: have nation-wide involvement of all interested parties, involve clear goal setting for each program monitored, and have agreement on the best performance indicators of reaching goals.²

The system should be simple. It should involve few performance indicators, and have clear and consistent rules for computation that can easily be done throughout the nation. While usually gross outcomes are measured, performance indicators should be stated in relative terms to facilitate cross region and cross program comparisons. The process of creating the system should be inclusive so as to achieve a consensus and sense of ownership that will promote professionalism and use of the system. The performance indicators system should be viewed as a changeable organic process that benefits from regular periodic refinement.

²Auer (1996) provides examples of performance indicators for employment policy among western European countries O'Leary (1995) provides examples from eastern Europe.

A main appeal of performance monitoring is that it provides a basis for a useful management information system for program operations. Focus on outcomes also promotes a culture of cost effectiveness and professionalism among employment service staff. Usually such a system involves follow-up surveys so that survey skills are established. The information system and survey skills combine to provide an excellent foundation for further evaluation studies.

Problems can arise in such a system. In particular where surveys are required response rates are always uneven across regions. Furthermore, when high performance is required there is incentive for data tampering at the local level. Finally high performance also means that *creaming* in program assignment is a distinct possibility. Resulting in wasted social resources.

Monitoring performance in Hungary

To provide an example of the results of performance measurement we draw on the experience of Hungary where nationwide performance measurement began in January of 1994. Hungary is currently in the process of developing a relational data base management system for administration of labor market support programs. Therefore, during the first five years of measurement performance indicators have been computed by hand on data aggregated at the county level for the 20 counties in Hungary. Table 1 reports on this evidence.

To present a simple summary of the performance measurement results we focus on a single type of performance measure that is available across the main active labor market programs. The measure summarizes the rate of reemployment. It is the percent employed after participation in each of the listed active labor market programs. Table 1 lists results for three types of retraining, self-employment, wage subsidy and public service employment (PSE) for the years 1994 through 1998.

Table 1. An example of performance measurement in Hungary. Percent employed at follow-up after various ALMPs, 1994-1998.

ALMP	1994	1995	1996	1997	1998
Group Retraining (A12)	44.9	36.1	44.5	46.3	46.8
Individual Retraining (A22)	58.5	42.2	51.9	51.1	51.5
Retraining Employed (A32)	82.2	93.6	92.8	90.4	94.7
Self-employment (B2)	91.9	90.6	90.2	88.1	91.7
Wage Subsidy (C2)	71.1	71.4	70.1	66.3	59.1
PSE (D2)	3.5	1.3	1.3	1.9	1.9

For people participating in ALMPs, the results indicate considerable stability in reemployment rates across the years following each separate program. For each program the reemployment rate fluctuated in a narrow range during the five year period. Also, the relative success in gaining regular non-subsidized employment across the different programs is quite stable over the years. The ordering from high to low of reemployment rates remains the same in every year.

In reviewing the results it must be remembered that follow-up was usually done 3 months after participants left the program. While this may be reasonable for retraining of the unemployed, it is quite soon for self-employment participants, and is probably too soon for the retraining of the employed and wage subsidy recipients.

Uses of performance monitoring results

A performance measurement system developed for ALMPs will have many uses for management, but the emphasis in these uses should always be on positive incentives rather than punitive action. Generally there are five principal uses:

- (1) To preserve decentralized decision making about allocation of funds to various programs and service providers.
- (2) To promote superior performance by counties, local offices, and service providers through positive incentives.
- (3) To help identify and correct poor performance through technical assistance and/or sanctions.
- (4) To contribute information on performance to the funding allocation process used by the tri-partite National Labor Market Committee to allocate funds to the counties.
- (5) To ensure compliance with legal requirements of programs.

The outcome found in Hungary that reemployment from individual retraining runs about 5 percentage points higher than for group retraining has influenced training program operations. Attention was drawn to alternative types of training by the performance monitoring system. In addition to reemployment rates, cost of reemployment was also a factor. It was observed that Budapest had used individual retraining at the near complete exclusion of group retraining, and that the cost per participant reemployed was very high. Budapest was encouraged to reconsider practices. Other areas also changed practice based on guidance from the National Labor Center.

Caveats on performance indicators

Since regions within a country vary in their economic strength, before using data on program performance in deciding budget allocation it is important to account for variations in the difficulty of finding reemployment. Consequently, an adjustment methodology for performance indicators is necessary. In addition to accounting for regional differences in reemployment prospects, the adjustment methodology may also provide an easy way to discourage "creaming" and ensure appropriate targeting of reemployment services.

Creaming refers to the practice of program administrators selecting the most qualified candidates for program participation so as to increase measured program success. The analogy is to milk where the richest part, the cream, floats to the top and can be skimmed off. Creaming is an issue in operating labor market programs because if only the most able people get reemployment assistance, then the benefit to society of the programs is not as great as it might be otherwise. Highly qualified program entrants have a good chance of becoming reemployed even without the services offered in the program, while for less qualified applicants the program services might be the only realistic path to employment.³

An appropriately designed adjustment methodology is an essential component of a performance management system. In addition to providing a level playing field for comparison of inter-regional performance, and a means for discouraging creaming by program managers, an adjustment methodology can be used to encourage targeting of services to those who have particular difficulty in gaining reemployment, such as: the long term unemployed, those with low levels of formal education, and persons with physical handicaps.⁴ Annex 2 to this report presents a technical discussion of how to develop an adjustment methodology for performance indicators. A properly designed adjustment methodology can be used to create incentives to prevent creaming and can provide for a fair assessment of program performance across regions.

³Evidence of creaming in assignment to training in Hungary is provided by Godfrey, Lazar and O'Leary (1993) and by O'Leary (1997).

⁴O'Leary (1996) provides a simple example of how to develop and apply an adjustment methodology for employment

4. Net Impact Estimation

The essential distinction of net impact estimation is that outcomes of program participants are judged relative to an appropriate comparison group. For employment programs this means that those personal characteristics that enable labor market success are roughly the same in the two groups. Appropriate comparison group specification can be achieved by proper sample selection or through statistical means. That is, either by classical field experiments or by quasi-experimental statistical methods. By taking care in estimation the process yields net rather than gross impacts. Net impacts are the proper indicator for judging the additional social value of an employment program.

Classically designed experiments are the ideal for net impact evaluation. If random assignment is achieved, modeling of behavior and complex econometric methods are not needed to obtain estimates of the net impact of a program. With large samples randomly assigned to treatment and control groups, observable and unobservable characteristics of the two groups should not differ on average, so that any difference in outcomes may be attributed to the program. Program impact may be measured as the simple difference between the means of the samples of program participants and of control group members on measures of outcomes. Because this process is easy to understand, simple unadjusted net impact estimates from field experiments are usually very influential for the purpose of guiding policy.⁵

Naturally, field experiments are not without potential problems. The first type of problems are called internal validity problems. These include errors in conducting random assignment to treatment and control groups, and inconsistent experimental conditions. The first of these can lead to lack of homogeneity across groups, the second means that the same treatment was not applied in all cases. The second type of pitfalls are called external validity problems. Time horizon effects can occur when treatment subjects understand that the experimental service is only temporary rather than permanent. Learning effects can take place within a community during the course of an experiment whereby the first enrollees act differently from those enrolled some time after the experiment begins. Hawthorne effects are responses to treatments not due to the content of service, but simply due to the special attention. Displacement effects, which may be the most critical external validity concern, occur when treatment subjects improve their outcome at the expense of others who are not part of the experiment.

When there is non-random assignment to either a program participant group or the comparison group, then statistical methods of correction must be used to offset the selection bias in order to properly estimate the net impact of a program.⁶

Recent surveys of microeconomic evaluations of employment programs conducted by Fay (1996) for OECD member countries and by Meager and Evans (1998) for a selected group of countries emphasize the importance of accounting for *deadweight loss* and *displacement* effects when measuring the impact of the program. With a mixed bag of findings which reveal that the net impact of different Employment programs varies widely from one population subgroup to another, the authors of both surveys argued that *targeting* of services is crucial to maximizing the social dividend from public expenditure on employment programs.⁷

programs.

⁵For examples of employment programs evaluated using a classically designed field experiment, see Decker and O'Leary (1995).

⁶Such methods are called *quasi-experimental* because they attempt to mimic statistically the ideal of a true experiment based on random trials (Fay, 1996).

⁷That is for the following reasons. When an unemployed person participates in an Employment program which does not improve his/her chance of re-employment, there is a *deadweight loss* to society for the expenditure incurred. If a program manager practices creaming in selecting participants for Employment programs such that the people supported would have secured employment without the assistance, then a deadweight loss also results. When an Employment program participant gains re-employment at the direct expense of an otherwise similar job-seeker, then *displacement* has occurred.

It is crucial to account for displacement and substitution effects when assessing the net social benefits of public programs. However, these factors are irrelevant at the individual level and very difficult to measure at the social level. An evaluation design using a comparison group automatically accounts for possible deadweight loss by comparing employment program participants with otherwise similar non-participants. A subgroup analysis of net impact provides a basis for targeting employment programs.

Quasi-experimental evaluations are often done because they are much cheaper and can be done more quickly than classical experiments. They can often be done with existing administrative data that further reduces evaluation costs. This is often the case when there is a "natural experiment," which is an opportunity presented by a policy change or an economic event. The main problems with quasi-experimental net impact evaluations concern adequately dealing with the problem of selection bias. This is a thorny issue that often requires complex statistical techniques to properly address. Such statistical complexity diminishes the policy value of the findings. Also, like experiment based net impact evaluations, the estimates only provide a snapshot photo at a point in time. This is distinct from the monitoring approach that gives consistent information covering a wide geographic area regularly over time.

To show the range of policy relevant outcomes that can be studied, we now examine net impacts of ALMPs in Hungary on employment, earnings, and receipt of unemployment compensation. Table 2 summarizes results for five separate outcomes.

Table 2.
Net impact of ALMPs on employment, earnings, and unemployment compensation in Hungary

	EMPLOYED ¹	EMPLNOW ²	EARNNOW ³	UCMONTHS ⁴	UCPAY ⁵
Hungary					
Individual retraining	0.11**	0.09**	7	-0.68**	-43 **
Group retraining	0.09**	0.07**	5 **	-0.50**	-27 **
Public service employment	-0.26**	-0.21**	9 **	-0.19	.9 **
Wage subsidy	-0.11**	-0.06**	-6	0.04**	7
Self-employment	0.14	0.16	-26	-1.64**	-120

** Statistically significant at the 95 per cent level in a two-tailed test

¹ Ever re-employed in an unsubsidized job or in self-employment

² Employed in an unsubsidized job or in self-employment on the survey date

³ Average monthly earnings from the current job on the survey date (US\$)

⁴ Months of unemployment compensation collected since January 1996

⁵ Amount of unemployment compensation collected since January 1996, in US\$ at exchange rate of US\$1.00 = 175.75 Hungarian forints on 1 April 1997, approximately the survey date

Source: O'Leary, Kołodziejczyk, and Lazar (1998).

Retraining and self-employment increase employment rates. Earnings appear to be boosted by group retraining and PSE, with no measurable effect from other programs. For unemployment compensation, retraining appears to reduce receipt by about one-half month while there appears to be even larger savings from self-employment assistance of more than 1.5 months.

When an employer, either government or private, receives a subsidy to hire a worker who would otherwise have been hired anyway, then *substitution* of Employment program financing for other intended spending has occurred. Johnson and Tomola (1977) provide a clear example of how to estimate the employment effects of fiscal substitution in direct job creation programs. They maintain that the degree of substitution increases as a program matures.

Subgroup analysis of net impacts

There are at least two reasons to examine program impacts by population subgroup. One is to provide information to policy makers who may consider targeting ALMPs to certain groups like those without a specialization or older unemployed persons. Another is to identify any possible biases in the effects--a program that benefits only one gender or certain education level groups may not be considered good policy even if it is cost effective. Table 3 presents the qualitative results of an analysis of ALMP impacts on important subgroups.

For neither individual nor group retraining were there marked differences by sex, age, education or occupational group. Subgroup analysis of participation in PSE indicated that participation was less likely to harm the re-employment chances of women, persons aged 45 years or over, and the better-educated. The greatest benefit of the wage subsidy was felt by participants in areas of moderate unemployment. However, impacts of wage subsidies did not vary appreciably by sex, age, or prior occupational group. Self-employment assistance boosted re-employment rates most among participants aged 45 years and over, and those in areas of high unemployment.

Table 3. Summary of Subgroup Net Impact Analysis

Characteristic	Retraining	Public Service Employment	Wage Subsidies	Self-employment
Gender		Worse for males		
Age				Best for older persons
Education		Worse for the less educated		
Occupation				
Unemployment Duration				
Unemployment Rate			Best where unemployment is moderate	Best where unemployment is high

Impacts of program features

Since ALMPs provided to unemployed job seekers are not homogenous, it is useful to investigate if variations in different observable dimensions of programs yields different impacts on the outcome measures for employment and earnings. Again drawing on the evaluation done in Hungary, Table 4 presents a qualitative summary of net impact estimates of various program features. As for the above discussion of subgroup impact estimates, all estimates presented apply to the outcome "currently employed in a non-subsidized job or self-employment" (EMPLNOW).

For both individual and group retraining in Hungary though the impact on employment was not significantly different from that for the complementary group it was greater for those who had contributed personally to the direct cost of individual retraining, for those in retraining for three months or less, and for those on courses involving 20 or fewer hours per week.

Participation in PSE work, which involves unskilled manual labor, appeared to be the greatest obstacle to future employment in a normal unsubsidized job or in self-employment, whereas non-manual and skilled manual work constituted the least impediment. There was no significant difference by the industry of activity

in which PSE occurred, but the reintegration of ALMP participants into the normal work force appeared to be more successful for PSE in service employment than in other industries.

Table 4. Summary of Program Feature Net Impact Analysis

Feature	Retraining	Public Service Employment	Wage Subsidies	Self-employment
Share in costs	Better with contribution (double but not stat. signif.)			
Duration of ALMP	3 to 12 months			
Organized by	Not district retraining center 20+ hrs/w			
Level of skill		Manual unskilled is worst	Outside of construction and services	Outside of services
Industry				
Sole proprietor vs. partnership				

The skill level required in the job that was given a wage subsidy had no significant effect on the employment outcome. Judged by whether participants were in unsubsidized work or self-employment on the survey date, those whose subsidized work had been in construction or the services reaped the least advantage from the wage subsidy scheme.

Self-employment assistance recipients who pursued activities in services industry were the least likely to experience lasting employment effects. There was not a significant difference in employment outcomes for those who started sole versus partnership business activities.

Not a part of the program feature analysis, but an important indicator of program impact secondary employment effects of self employment assistance in Hungary was also investigated. It was found that on the survey date 17.6 percent of those receiving assistance had hired at least one other worker. Indeed, one successful recipient claimed to have hired 12 workers. The mean number of workers employed by those who did recruit was 1.75, and the mean hired among all assistance recipients was 0.31. About half of hires had previously been unemployed.

Every ALMP has a variety of features and important outcomes. The presentation in this and the previous sections of chapter 4 are not meant to be an exhaustive enumeration of all outcomes or evaluation approaches possible. The analysis and examples were intended to be suggestive of the potential value of the net impact approach to program evaluation.

Uses of net impact estimates

In the previous chapter the point was made that net impact analysis is a supplement to performance monitoring. The latter being mainly a management tool and the former useful for policy development. There is much more overlap in use of the two sets of results than that division suggested. This section briefly discusses four uses of net impact estimates of ALMPs. The four are: (1) policy formulation, (2) targeting, (3) program management, and (4) accountability. In our exposition we hope to make clear the added value which net impact analysis provides to administrators, users, and decision makers of employment programs.

Policy formulation decisions concerning questions of whether to continue, expand, curtail, or cancel government employment programs should be well informed with objective information. Policy makers within labor ministries, national labor centers, and national legislatures often require information about the return on government spending--the return on investment. The net benefits for programs may be assessed from different perspectives: society, individuals, government, and programs. Such a measurement requires estimates of the incremental value of programs in cost-benefit analysis.

To improve the overall cost effectiveness of programs and to increase the value to customers, it is useful to know what programs yield the greatest benefit for different clients. Subgroup analysis of program net impacts can provide exactly the information needed to do informed targeting of reemployment services. Such information is available no other way.

As emphasized in chapter 3, performance indicators systems are a valuable tool for program management. Such systems track gross outcomes and by themselves have no mechanism for establishing what is an adequate or superior level of performance. Net impact analysis can provide baseline standards for gross outcome performance monitoring systems. Such standards can be used with an adjustment methodology system to set reasonable targets for regions on separate employment programs.

Spending of public funds for social improvement requires public accountability. Periodic reports to parliament, prime ministers, and voters are crucial for continuance. Either over the short term or the long term all spending programs are discretionary. Good programs only survive and flourish when credible evidence of value can be objectively demonstrated. The methods reviewed in this chapter provide strategies for determining which public efforts to promote employment provide the greatest value.

5. Conclusion

When the net benefits of an ALMP are being evaluated, it must be made clear whether the program is being assessed from a social, governmental, program or individual perspective. It is impossible properly to measure every detailed factor bearing on such computations. However, it is possible to measure the main elements for such computations and net program impacts are central.

It should be recognized, however, that the decision to implement, continue, curtail or cancel programs for labor market support also has a political dimension. During a period of dramatic change in conditions of employment security, there can be an irresistible imperative for ALMPs. In such times, the rules for return on investment cannot be blindly applied without regard for social stability, however difficult that may be to quantify.

Aside from their net impact, ALMPs have a direct effect of easing labor market tensions because of the simple fact that those taking part in ALMPs are not counted as unemployed during their participation. In Hungary during the mid-1990s for example, participation in ALMPs reduced measured unemployment by 2 percentage points below what it would have been otherwise. Furthermore, while ALMP participation does not always immediately result in stable re-employment, for individual program participants the experience serves at least to interrupt an otherwise continuous spell of unemployment.

In times when unemployment is high and the demand for labor is low, we should not expect uniformly high net impacts from ALMPs. The examples from Hungary reported in this paper highlight the importance of carefully assessing active labor programs so that public funds can be utilized as efficiently as possible while pursuing the social goal of returning the unemployed to gainful work.

REFERENCES

- Auer, Peter. 1996. "Monitoring of labour market policy in EU Member states", in *Policies: inforMISEP* (Berlin), Number 53 (Spring), pp. 24-31.
- Decker, Paul T.; O'Leary, Christopher J. 1995. "Evaluating pooled evidence from the reemployment bonus experiments", in *Journal of Human Resources* (Madison, WI), Volume 30, Number 3 (Summer), pp. 534-550.
- Fay, Robert G. 1996. *Enhancing the effectiveness of active labour market policies: Evidence from program evaluations in OECD countries*. Labour Market and Social Policy Occasional Papers No. 18. Paris, Organisation for Economic Co-operation and Development.
- Godfrey, Martin; Lazar, Gyorgy; O'Leary, Christopher. 1993. *Report on a survey of unemployment and active labour market programs in Hungary*. ILO/Japan Project on Employment Policies for Transition in Hungary. Budapest, Ministry of Labour, October.
- Johnson, George E.; Tomola, James D. 1977. "The fiscal substitution effect of alternative approaches to public service employment policy," *Journal of Human Resources* (Madison, WI), Volume 12, Number 1 (Winter), pp. 1-26.
- Meager, Nigel; with Evans, Ceri. 1998. *The evaluation of active labour market measures for the long-term unemployed*. Employment and Training Papers No. 16. Geneva: ILO.
- O'Leary, Christopher J. 1995. "Performance indicators: A management tool for active labour programs in Hungary and Poland", in *International Labour Review* (Geneva), Volume 134, Number 6, pp. 729-751.
- O'Leary, Christopher J. 1996. *Methods for performance-based management of active labor programs in Hungary: An adjustment methodology for performance indicators and a proposal for budget allocation*. Kalamazoo, MI, W.E. Upjohn Institute for Employment Research.
- O'Leary, Christopher J. 1997. "A net impact analysis of active labour programs in Hungary", in *Economics of Transition* (Oxford), Volume 5, Number 2 (October), pp. 453-484 .
- O'Leary, Christopher J., Piotr Kolodziejczyk, and Gyorgy Lazar. 1998. "The Net Impact of Active Labour Programs in Hungary and Poland", in *International Labour Review* (Geneva), Volume 137, Number 3, pp. 321-346.
- O'Leary, Christopher J., Alena Nesporova, and Alexander Samorodov. 2001. *Manual on Evaluation of Labour Market Policies in Transition Economies*. Geneva: ILO.