1-1-1999

A Literature Review on Methodology for Measuring and Monitoring Net Impacts of Employment Benefits and Support Measures

Christopher J. O'Leary
W.E. Upjohn Institute, o'leary@upjohn.org

A Literature Review on Methodology for Measuring and Monitoring Net Impacts of Employment Benefits and Support Measures

March 1999

Prepared by:

Christopher J. O'Leary
W. E. Upjohn Institute for Employment Research
300 South Westnedge Avenue
Kalamazoo, Michigan 49007 USA
(CAC Vendor No.: 0V4351)
Tel: (616) 343-5541
Fax: (616) 343-3308

Prepared for:

Evaluation and Data Development
Human Resources Development Canada
112 Kent Street
22nd Floor, Tower "B"
Ottawa, Ontario K1A 0S5 Canada

Ron Rocheleau, Project Officer
Tel: (613) 957-8191
Fax: (613) 954-0134

Prepared under Consulting and Audit Canada Contract No.: 330-9062 Project No.: 330-0722-16.

A Literature Review on Methodology for Measuring and Monitoring Net Impacts of Employment Benefits and Support Measures

TABLE OF CONTENTS

A Literature Review on Methodology for Measuring and Monitoring Net Impacts of Employment Benefits and Support Measures

Abstract

This paper reviews the literature on evaluation of government operated programs to provide temporary income support and to promote reemployment for unemployed job seekers. The paper is a synopsis of the international literature spanning various evaluation techniques. The main aim is to identify the best methodology for measuring and monitoring net impacts of Employment Benefits and Support Measures (EBSMs) in Canada. Sections of the review are organized around the sources of literature. The first source is the academic literature, which provides a theoretical overview of alternative approaches to measuring and monitoring program impacts. The issue of how an ideal net impact monitoring system should work is also explored. The next source of literature is Canadian and foreign government reports. The literature review provides a context to assess the Canadian experience and future directions. The Upjohn Institute team has a high degree of foreign experience on similar topics. This was an asset in preparing the literature review since the government and other non-academic literature is typically difficult to obtain by the normal channels.

A Literature Review on Methodology for Measuring and Monitoring Net Impacts of Employment Benefits and Support Measures

A. AIMS OF THIS LITERATURE REVIEW

This paper reviews the literature on evaluation of government operated programs to provide temporary income support and to promote reemployment for unemployed job seekers. The paper is a synopsis of the international literature spanning various evaluation techniques. The main aim is to identify the best methodology for measuring and monitoring net impacts of Employment Benefits and Support Measures (EBSMs) in Canada.

Sections of the review are organized around the sources of literature. The first source is the academic literature, which provides a theoretical overview of alternative approaches to measuring and monitoring program impacts. The issue of how an ideal net impact monitoring system should work is also explored. The second source of literature is Canadian and foreign government reports.

The literature review provides a context to assess the Canadian experience and future directions. The Upjohn Institute team has a high degree of foreign experience on similar topics. This was an asset in preparing the literature review since the government and other non-academic literature is typically difficult to obtain by the normal channels.

B. AN OVERVIEW OF EVALUATION METHODS

B.1 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 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 easily misguide program management and policy decisions. Net impacts compare 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 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 examining the rate of reemployment. Suppose that the rage 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. 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.

B.2 Evaluation Techniques

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*, while it is usually done by a cheaper and quicker *quasi-experimental method* which relies on statistical methods to mimic an experiment.

B.3 Issues in 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. The system should be simple, involve few performance indicators, have clear and consistent rules for computation which can easily be done throughout the nation. While usually gross outcome measures, 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 which will promote professionalism and use of the

system. The performance indicators system should be viewed as a changeable organic process which benefits from regular periodic refinement.

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 and 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 is wasted social resources.

B.4 Issues in 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 which 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.

B.4.1 Classically Designed Experiments

Classically designed experiments are the ideal for net impact evaluation. If random assignment is achieved, modeling of behaviour 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.[1]

Naturally, field experiments are not without potential problems. The first type of problems are called internal validity problems. These include errors in conducting random

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

assignment to treatment and control groups, and inconsistent experimental conditions. The first problem 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 pitfall 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.

B.4.2 Quasi-experimental Econometric Evaluations

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.[2]

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.[3]

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

---

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

[3] That 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 practises 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. 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.

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 which 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 is adequately dealing with the problem of selection bias. This is a thorny issue which 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 which gives consistent information covering a wide geographic area regularly over time.

C. MONITORING GROSS PROGRAM OUTCOMES

C.1 An Overview of Monitoring Gross Program Outcomes

The monitoring of program outcomes is the basis of performance management systems. In principle such systems can support decentralized decision making in employment policy while promoting superior performance through positive incentives. Performance indicators which are the core of each of these systems are concise measures of program performance that can be easily computed and tracked on a regular basis.

To operationalize performance management systems information from the monitoring of program effectiveness is used by computerized management information systems which combine data from administrative records with evidence on participant outcomes. Such evidence may come from follow-up surveys of program participants and service providers to present an informative view of program effectiveness. The performance indicators can become a natural tool for managers, allowing informed and objective decisions at the local, regional and national levels.

To develop good performance indicators the goals of employment programs must be clearly understood. Performance indicators should be set to guide operations toward program goals, but the most important principle governing the development of performance indicators is that program *outcomes* rather than process be emphasized.

The monitoring of program outcomes must not impose an excessive administrative burden on local and regional employment offices where the first priority must be service to clients. The list of performance indicators should be relatively short for any particular program, and the associated follow-up surveys should ask the minimum possible number of questions. By limiting performance measurement to a small number of indicators, any required follow-up surveys may also remain simple. This will increase the reliability of data gathered, increase the response rate, and increase the likelihood that the system will survive over time thereby yielding valuable information on how programs perform over time.

A basic objective of evaluating employment programs is to compare their relative cost effectiveness in promoting employment. They should be constructed so as to measure output per unit of input. The ultimate success of any employment program occurs when a program participant either gains regular employment or avoids unemployment with the assistance provided. The average expenditure to achieve this result is the basic measure for comparing cost effectiveness across programs.

In an effort to promote an optimal mix of employment programs, performance indicators can be incorporated directly into the annual planning and budget allocation processes. The incentive to achieve a high level of effectiveness will be strengthened if even a small part of the annual budget allocation depends on measured performance.

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.

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]

The over-riding goals of employment programs are to achieve reemployment of unemployed persons, and prevent unemployment for workers at risk of job loss. Two categories of performance indicator measure the success in achieving this goal: rate of reemployment, and cost of reemployment. A corollary aim of employment programs is to smooth the transition between jobs. The main category of performance indicator measuring for this goal is support cost. Beyond these general goals, there are a variety of other goals specific to the various employment programs. This diversity of goals reflects the impossibility of serving all needs with a single program.

Another part of the strategy in developing performance indicators is to specify them so that comparisons across programs are possible. Certain of the performance indicators across programs should be similar enough to allow this. The most comparable measure across programs falls under the category cost of reemployment. In the performance indicators, this is usually based on measurement of employment at follow-up. Other categories of performance indicators such as the rate of reemployment, and the support cost, can also allow for comparison across programs, but the performance indicators formulae for measurement across programs are less similar due to the differences in program design.

---

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

Possible uses of performance indicators in managing employment programs are:

(1) To preserve decentralized decision making about the allocation of funds to various programs and service providers.

(2) To promote superior performance by regions, 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 budget allocation process.

(5) To ensure compliance with legal requirements of programs.

The emphasis among all these uses is on positive incentives rather than punitive action.

C.2 Academic Literature on Monitoring Gross Program Outcomes

Simpson (1990) gave an overview of evaluating gross outcomes of social programs. Osborne and Gaebler (1992) provide documentation of innumerable cases where performance indicators are used by state and local governmental units in the United States. Auer (1996) documents the use of performance measurement systems for employment programs by countries in the European Union. The OECD (1994) provided a guide for program managers on how to use such a system. Following the lead of countries in Western Europe (Auer, 1996), the Central European transition economies of Hungary and Poland have adopted performance indicators systems as a means of monitoring active labor program effects. O'Leary (1995) provides a summary of their system. O'Leary (1993) suggested approaches for avoiding creaming in program assignment for performance managed employment programs.

In specifying performance indicators for employment programs it is important that the intermediate goals which result from the performance indicators are consistent with the broad objectives of securing appropriate regular employment and maintaining adequate income support. High performance as measured by the performance indicators should not have unintended negative side effects. The issue of incentive compatibility of performance indicators with larger aims has received quite extensive attention in the research literature; important papers are: Barnow (1992), Dickinson et al. (1988), and Singer (1986). Even after considering the problems of incentive compatibility and creaming, Carolyn Heinrich (1995, p. 66) concludes that:

> when performance standards are carefully developed to align incentives with program goals, policymakers should be able to facilitate the kinds of social program outcomes and impacts they want without having to formulate detailed

rules and direct management controls to guide program administration and service delivery.

C.3 Government Literature on Monitoring Gross Program Outcomes

Performance indicators are a widely accepted method for managing public programs. Green and Aaronson (1992) discuss how performance indicators are used in managing training and education programs in 39 programs which are administered by 7 departments of the United States federal government.

In 1993 the federal government of the United States enacted the Government Performance and Results Act (GPRA). So as "to improve the confidence of the American people in the capability of the Federal Government, by systematically holding Federal agencies accountable for achieving program results." (U.S. Congress; 1998, p. 32564). Under the Act all federal agencies are to operate management systems based on performance measures.

Extensive systems of performance indicators are also used in the United Kingdom with some measures of performance being regularly reported in the *Employment Gazette.* The U.S. General Accounting Office (1998) provided a constructive assessment of the performance management systems used for employment programs in the United States.

Dickinson et al. (1988) evaluated the effects of performance standards on clients, services and costs under the Job Training Partnership Act. West (1992) explains how adjustment methods for performance standards were developed.

Following the lead of countries in Western Europe (Auer, 1996), the Central European transition economies of Hungary and Poland have adopted performance indicators systems as a means of monitoring active labor program effects. O'Leary (1996, 1994) in government reports to Hungary provides a summary of their system. The first version of this system is given in O'Leary (1990). O'Leary and Targowski (1993) in a government report to Poland provides a summary of their system.

D. LITERATURE ON NET IMPACT ESTIMATION

D.1 An Overview of Net Impact Estimation

Since there is a possibility of selection bias in assigning registered unemployed persons to active labour programs (ALPs), special care must be taken in evaluating the net impacts of ALPs on labour market success. It is useful to have knowledge of three program impact estimation methods: (1) simple unadjusted difference in means, (2) difference in means using a matched pairs comparison group, and (3) regression adjusted impact estimates. The following provides a brief description of each of these procedures. Also given is a concise statement of the subgroup impact estimation methodology, and some other relevant procedures.

Unadjusted Impact Estimates

When random assignment has been achieved, modelling of behaviour and complex econometric methods are not needed to estimate reliable net program impacts. 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 exposure to the program. Program impacts may be computed as the simple difference between means of the samples of program participants and control group members on outcome measures of interest, or:

$$(1) \qquad E(y_p) - E(y_c),$$

where E is the expectation operator yielding means of the random variables, y is an outcome of interest, and the index p denotes the sample of program participants while c denotes the comparison sample. Tests of significance are done using t-statistics.

In terms of clearly guiding policy, simple unadjusted net impact estimates based on random trials are usually the most influential because they are easy to understand. This is the main appeal of program evaluation done using a classically designed experiment involving random assignment.[5] The result of the computation stated in equation (1) is equivalent to the slope coefficient estimated by ordinary least squares (OLS) applied to a simple bivariate regression model. That is, program impacts can be estimated by running the OLS model:

$$(2) \qquad y_i = a_0 + a_1 P_i + u_i,$$

on a pooled sample of comparison group members and program participants, where y is the outcome of interest, a_1 is the impact of the program on the outcome for the ALP participants, a_0 is the mean value of the outcome for comparison group members, P is a dummy variable with a

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

value of 1 for active labour program (ALP) participants and 0 otherwise, u_i is a normally distributed mean zero error term, and I is an index denoting individuals in either the participant or comparison group samples. Tests for significance of program impacts are simply t-tests on the parameter a_1.

Impact Estimates Using a Matched Pairs Comparison Group

When participant group and comparison group members differ significantly in terms of observable characteristics, it would not be surprising to observe different labour market success across program participant and comparison groups even in the absence of ALPs. To put the assessment of ALPs on an even footing, a separate comparison group for each sample of ALP participants may be formed using a matched pairs methodology.[6]

If it is representative of the universe of unemployed, the sampling frame for the comparison group can be randomly selected from the unemployment register. Matched pairs comparison groups can then be formed by comparing persons in the ALP participant samples with those in the full comparison group using the standardized Mahalanobis distance measure:

$$(3) \qquad d_{pc} = \mathrm{Sum}_k(Z_{pk} - Z_{ck})^2$$

where, the index p represents observations in an ALP participant sample and the index c represents observations from the comparison group, the index k runs over the n exogenous characteristics on which the observations are matched, and Z represents the standardized value of a characteristic where the mean and standard deviation of the characteristic is computed on the pooled sample of the comparison group sampling frame and the participants in the relevant ALP.

Using this distance measure, separate matched pairs comparison groups were selected for each ALP. The person with the smallest d_{pc} from the full comparison group sampling frame was selected for inclusion in the matched pairs comparison group, with ties being resolved randomly and each person in the ALP sample being compared to all those in the full comparison group sampling frame.[7]

After forming the matched pairs comparison groups, program impact estimates were computed using a simple difference of means, with significance of impacts being judged by t-tests. It should be noted that because a single observation from the comparison sample may be chosen more than once for the synthetic comparison group, the estimated standard error, computed in the usual way, for this group will be reduced. Therefore standard t-tests based on

[6] See Fraker and Maynard (1987) for an interesting review and application of comparison group designs for evaluating employment-related programs.

[7] That is, sampling was done with replacement.

matched pairs data depend on lower bound standard error estimates which give the upper bound on the possible statistical significance.

Regression Adjusted Impact Estimates

Multivariate regression analysis is a natural method for assessing the net impact of program participation on labour market success when observable characteristics of participant and comparison group members are dramatically different. This method involves a simple extension of equation (2). In such cases, estimation of the model:

$$(4) \qquad y_i = a_0 + a_1P_i + b_1X_{1i} + b_2X_{2i} + ...+ b_nX_{ni} + u_i,$$

by OLS on the pooled sample yields net program impact estimates.[8] In equation (4) y is the outcome of interest, a_0 is the mean value of the outcome for comparison group members evaluated at the mean of all observable characteristics included in the regression, P is a dummy variable with a value of 1 for program participation and 0 otherwise, a_1 is the impact of the program on the outcome for the program participants evaluated at the mean of all observable characteristics, X_1 to X_n are observable characteristics measured as deviations from their mean values, u_i is a normally distributed mean zero error term, and I is an index denoting individuals in either the participant or comparison group samples.[9]

This method yields net program impacts adjusted for observable characteristics.[10] The estimates are called net because the comparison and program participant groups are statistically adjusted so as to remove heterogeneity across the samples. That is, the only remaining factor contributing to a difference in the outcome measure is exposure to the program treatment. The estimation methodology nets out all other observable factors affecting the outcome.

---

[8] In this report, since the main dependent variable of interest--in a normal job--is binary, the regression model predicts the probability of reemployment. The OLS estimation is a linear probability model, which may yield biased estimates. OLS estimates may be biased since the range of variation in the dependent variable is constrained to the zero-one interval. Maddala (1982, Chapter 1) suggests using the logit estimator in such cases. Bias is usually most severe when the bulk of probability clusters at one or other extreme of the zero-one interval Since reemployment probabilities for the ALP and comparison groups generally range from about 40 to 60 percent, the limited range of the dependent variable is not a likely source of severe bias in estimating parameters by OLS.

[9] In this application the regression model is a statement of an analysis of covariance methodology, where X_1 to X_n are the covariates. Mohr (1992, pp. 83-87) discusses extending a regression model for program impacts to include control variables.

[10] The obvious next procedure to adjust for differences across samples is to account for differences in unobservable characteristics. The technique, which involves applying the methods of Heckman (1976), is problematic because instruments are usually not available to explain program participation independent of reemployment success.

Full Interaction Regression Adjusted Impact Estimates

A more general regression model for impact estimation which allows for variation in program effects by observable characteristics during estimation is called a full interaction regression model. Such a model is a direct generalization of equation (4). The model may be written:

(5) $y_i = a_0 + a_1P_i + b_1X_{1i} + b_2X_{2i} + ...+ b_nX_{ni} + c_1P_iX_{1i} + c_2P_iX_{2i} + ...+ c_nP_iX_{ni} + u_i,$

and can be estimated by OLS on the pooled sample to give net program impact estimates. In equation (5) the variables are the same as those defined for equation (4). However, for this generalized regression model the net program impact is computed as $a_1 + sum_k(c_kE(X_k))$, where $E(X_k)$ denotes the mean of characteristic X_k. Tests of confidence on these linear combinations of estimates may easily be performed as F-tests.

Subgroup Net Impact Estimation Methodology

For each separate ALP, subgroup treatment impacts were simultaneously estimated in a single regression model. The specification employed allows the treatment response for each subgroup to be estimated controlling for the influence of other subgroup characteristics. For example, the model allows estimation of treatment impacts associated with being female controlling for the fact that females are more likely to have more formal education and less likely to work in a blue collar occupation.

Suppressing subscripts and using matrix notation, the regression equation used to estimate subgroup net impact estimates can be written:

(6) $Y = a + PB + GC + GPD' + u$

where Y is the outcome measure, a is the intercept, B, C, and D, are conformable parameter vectors, P is the indicator of participation in an ALP, G is the matrix of dummy variables which code for membership in a subgroup, and u is a mean zero normally distributed random error term. Equation (6) specifies a complete one-way interaction model. It allows simultaneous estimation of all subgroup treatment impacts, but imposes linear restrictions on the estimates. Treatment impacts for a particular subgroup are computed as the sum of the parameter estimate on the product of the subgroup dummy variable and the treatment indicator plus the sum of parameter estimates on the product of subgroup dummy variables and the treatment indicator multiplied by their respective population shares. In each computation, parameter estimates for the complement to the subgroup of interest are omitted.

The subgroup impact estimates may be considered to be regression adjusted in the sense that each subgroup impact is estimated while simultaneously allowing impacts to vary across other subgroups considered.

Methodology for Estimation of Program Components

To estimate the impact of separate features of an ALP on outcomes of interest, new program variables are defined from the single program variable P_i such that the vectors for the new variables add up to the vector for the old variable. For example, if P_i has a value of 1 if participated in an ALP and 0 otherwise, to examine the separate impacts of the ALP operated by public and private enterprises on outcomes of interest we may define $P_{1i} = 1$ if participated in an ALP operated by a public enterprise and 0 otherwise, and $P_{2i} = 1$ if participated in an ALP operated by a private enterprise and 0 otherwise. Therefore $P_i = P_{1i} + P_{2i}$, and the separate impacts of the ALP run by public and private enterprises on outcomes of interest can be estimated by OLS regression applied to a simple model like:

(7) $y_i = b_0 + b_1P_{1i} + b_2P_{2i} + u_i.$

From this model the parameter estimate for b_1 is the impact of wage subsidy run by public enterprise on outcome of interest, while b_2 is the impact of wage subsidy run by private enterprise. The model of equation (7) can be applied to other partitions of the program experience, such as short and long duration participation, or to partitions which are more than two way, such as three industry groups for program operators.

Notice, that in this case the full set of indicator variables is included in the equation for OLS estimation. For this procedure the full set of program treatment indicators does not introduce singularity in estimation, because the program vectors include data on both program participants and comparison group members. Equation (7) also presumes that the participant and comparison groups are homogenous in observable characteristics. If this is not the case, control variables should be added to the specification as was shown in equation (4).

Method for Separating out Impacts of Multiple Programs

It is very possible that an individual may have participated in more than one ALP. In particular, it is a frequent occurrence that a participant in an ALP such as retraining or public service employment will also use the services of the employment service (ES) in an effort to gain reemployment. To estimate the impact of a single program when some in a sample being analyzed have used more than one program, a simple regression model may be used. Suppose that someone uses both an ALP and the ES, then a model like the following might be estimated:

(8) $y_i = a_0 + b_1ALP_i + b_2ES_i + b_3ALP_i * ES_i + c_1X_i + u_i,$

where ALP represents participation in an ALP, ES represents use of an ES service, X represents exogenous control variables, y is the outcome of interest, and u is a normally distributed mean zero error term. After estimating an equation of this form by OLS, the marginal effect of the ALP on y is estimated by the sum of $b_1 + b_3 * E(ES)$, where E is the expectation operator and E(ES) is the mean of the variable ES or the proportion of the sample which used the ES.

Similarly the marginal effect of the ES on y is estimated by the sum of $b_2 + b_3$ *E(ALP). ests of confidence on these sums of estimates may easily be performed as F-tests.

Methods for Analysis of the Timing of Response

To examine the impact of ALP participation on the time pattern of reemployment, conditional exit rates are examined for each month. The exit rate is computed by dividing the number of registered unemployed who left the register for reemployment in a given month by the number of claimants in the group at the start of that month. Letting h(t) denote the conditional exit rate in month t, and R_t the number of registered unemployed at the start of month t, then

(9) $h(t) = (R_t - R_{t+1})/R_t,$

is a conditional measure of a change in behaviour because it depends on the number who had yet to change their behaviour regarding the outcome at the start of each month (R_t). The expression h(t) is the popular Kaplan-Meier exit rate discussed thoroughly by Kiefer (1988). The number of registered unemployed at the start of each time period (R_t) is called the "risk set" because it is the number of job seekers "at risk" of changing behaviour in the subsequent month. Note that risk set in month t+1 equals the risk set in the previous month times one minus the exit rate for that month [$R_{t+1} = R_t (1 - h(t))$].

Sample Size Requirements for Power Tests of ALP Effects

Testing the difference between proportions is somewhat complicated by the fact that the sample sizes required for properly testing a given difference between proportions varies depending on whether the proportions are near zero or one (Cohen 1988, chapter 6). Specifically, the required sample sizes for testing the difference in proportions with adequate power depend on the effect size, h, which is the difference in the arcsin transformation of the proportions. That is, $f(p) = 2\arcsin\sqrt{p}$ and the effect size is $h = |f(p_p) - f(p_c)|$ for non-directional tests where p_p is the proportion employed among the ALP participant group and p_c is the proportion employed among the comparison group. For tests of $(p_p - p_c) = 0.05$ when p_p is around 0.5 then h = 0.1. To perform two tailed tests at the confidence level of 98 percent with a power of 80 percent and h = 0.1 the harmonic mean of the sample sizes should be at least 2,007 in size, where the harmonic mean, n', of the samples sizes is $n' = 2n_pn_c/(n_p + n_c)$. Lowering the confidence level to 90 percent lowers the sample size requirement to 1,237. When p_p is closer to either 0 or 1 the sample size requirements for similar tests [$(p_p - p_c) = 0.05$] are smaller.

D.2 Academic Literature on Net Impact Estimation

Net impact evaluation of government employment and training programs dates at least to Borus (1964) who evaluated impacts on employment and earnings of Manpower Demonstration and Training Act (MDTA) participants in the U.S. Heckman (1976) did papers identifying the effect of sample selection in program participation and exploring easy ways to correct for the

problem. Maddala (1986, pp. 260-267) summarized the essential methods developed by Heckman. An important treatment of sample selection bias was done by Ashenfelter (1978) who identified what has come to be known as the "Ashenfelter dip" in earnings prior to program participation. Ashenfelter and Card (1985) and others confirmed the existence of the phenomenon. The finding was identified more generally for displaced workers by Jacobson, LaLonde, and Sullivan (1993).

The Comprehensive Employment and Training Act (CETA) of 1973 was the first training program for which the U.S. Department of Labour developed a data base specifically intended for program evaluation (Leigh 1990, p. 10). It was called the Continuous Longitudinal Manpower Survey (CLMS) and contained data on program participants, data on comparison group members drawn from the national labour force survey (Current Population Survey), and earnings data for all subjects from national social insurance (Social Security) records. Despite the fact that CETA programs were targeted to low-income individuals while the labour force survey represented the nation, evaluation studies were greatly facilitated. Three main findings emerged from 11 major CETA evaluations (Leigh 1990, p. 11). First, there were no measurable employment or earnings impacts for men, however impacts for women were positive and significant. Second, on-the-job training is usually more effective than classroom training. Finally, the range of impact estimates was quite wide, despite the fact that all analysts used the same CMLS data. However, it was journalists rather than economists who brought the end to CETA. The pubic service employment (PSE) component of CETA became a national target for criticism when careless management of funds and enrollment of program ineligibles were widely reported. The associated problem of "fiscal substitution" where by local government agencies replaced regular staff with CETA PSE workers in order to conserve local tax payer money was estimated by Johnson and Tomola (1977) to increase with the maturity of the program. Problems in CETA PSE increased dramatically after funding for PSE was greatly expanded in 1977 as part of an expansionary federal fiscal policy. The CETA program ended without remorse in 1982 at its scheduled expiration date (Leigh 1990, p. 7). Even though states gained a role in administration of CETA vis-a-vis MDTA, in the end no states were advocates for continued authorization of CETA. By then, CETA had become a four letter word.

Johnson and Stromsdorfer (1990) summarize strategies for net impact evaluation of Job Training Partnership (JTPA) programs in the U.S. The Job Training Partnership Act (JTPA) of 1982 was the result of true ideological and partisan compromise. The bill was jointly sponsored by the liberal Democratic Senator Edward Kennedy and the conservative Republican Senator Daniel Quayle. Many features of the bill reflected the compromise. Evaluation was an integral part of the program which was said to be performance driven through a system of performance standards for participant reemployment rates and earnings. Also of note was the absence of anything remotely resembling PSE. The performance standards system allowed governors receiving federal JTPA training grants to structure incentive systems, and there by objectify the relationship to substate areas. Ostensibly removing politics from the funding process. Governors reserved some allocations for incentive rewards paid to areas achieving high levels of performance. The performance monitoring system changed training program management and

intergovernmental relations. It also complicated the net impact evaluation of programs by introducing the risk of *creaming* in program assignment.[11] That is, program managers might select mainly the most able applicants for participation. The result is high observed reemployment rates, however many of the selected program participants may already possess the skills and abilities to get reemployed themselves. By comparing their success to all unemployed, the positive impact on reemployment is high, but comparing their success to others with similar characteristics the program impacts may be much smaller.[12] To assure an objective net impact evaluation, Congress authorized a major national evaluation of JTPA based on methods of field experimentation with random assignment of subjects both to training and to comparison groups in 16 sites across the country. Orr et al. (1995, p. 109) report that training to economically disadvantaged adults resulted in 11 percent greater earnings for women and 6.7 percent greater earnings for men. For both genders the earnings gains were mainly due to increases in hours worked. There were positive net benefits to both men and women and the net benefit to society for both genders was just over $500 per participant (Orr et al.; 1995, p. 189).

For unemployment insurance (UI) in the U.S., since the work of Ehrenberg and Oaxaca (1976) it has been generally accepted that UI lengthens spells of insured unemployment beyond what they would be otherwise. Decker (1997) documented the range of estimates UI has on reemployment. He reported that a 10 percent increase in the rate at which UI benefits replace prior wages increases the duration of unemployment by between 0.5 and 1.5 weeks, and a 1 week increase in the potential duration of benefits increases unemployment duration by between 0.1 and 0.5 weeks. While these negative impact estimates are non-disputable, it is also possible that prolonged job search improves the quality of job matches which ultimately boosts worker productivity.

D.3 Government Literature on Net Impact Estimation

Since evaluations of unemployment compensation and retraining programs have usually been done under government contract most work in the academic literature has it's origins in government reports. Concerning unemployment insurance (UI) in the U.S., we are fortunate that the wisdom of Steve Wandner created the UI Occasional Paper series with in the U.S. Department of Labor. This series contains a wealth of knowledge and details about how to do evaluation of passive income support. Most U.S. evaluations of training programs remain largely unpublished technical reports from research institutes, universities and consulting firms.

---

[11] 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.

[12] An evaluation of retraining in Hungary found evidence of creaming in referral to services (O'Leary, 1997).

E. CONSIDERATIONS FOR MONITORING NET IMPACTS OF EBSMs

E.1 Conceptualizing the Problem

Efforts within HRDC to develop a high performance driven system for EBSMs based on continuous measurement of net/incremental impacts on outcomes is at the forefront of international practice for modern public management of government employment promotion programs. It is well known that when program managers are encouraged to achieve a high employment rate for program participants in a gross outcome based monitoring system, a phenomenon called *cream skimming* frequently results.[13] That is, program managers might select mainly the most able applicants for participation. The result is high observed reemployment rates, however many of the selected program participants may already possess the skills and abilities to get reemployed themselves. By comparing gross outcomes of participants to outcomes for all unemployed, the positive impact on reemployment is high, but comparing participant success to others with similar characteristics the program impacts may be much smaller.[14]

Since they are widely recorded on a continuous basis, the performance indicators results are useful for ongoing program management and planning. However, these indicators cannot inform policy makers about any added value which may be provided by programs. For such net impact analyses a comparison group design is needed. Net impact evaluations help policy makers decide which programs to expand, modify or delete as economic and political conditions change. If such evaluations were available on an ongoing basis they may serve as useful management tools supplanting gross outcome monitoring and removing several of the associated problems.

In terms of clearly guiding policy, simple unadjusted impact estimates are usually the most influential because they are easy to understand. This is the main appeal of program evaluation done using a classically designed experiment involving random assignment.[15] When random assignment has been achieved, modeling of behavior and complex econometric methods are not needed to estimate reliable net program impacts. 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 exposure to

---

[13] The analogy is to milk where the richest part, the cream, floats to the top and can be skimmed off. Cream skimming 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.

[14] An evaluation of retraining in Hungary found evidence of creaming in referral to services (O'Leary, 1997).

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

the program. Program impacts may be computed as the simple difference between means of the samples of program participants and control group members on outcome measures of interest.

When there is non-random assignment to either the program participant group or the comparison group from the population of unemployed job seekers then statistical methods of correction must be used to reveal the net impacts of programs.[16] That is, proper estimation of program net impacts involves correcting for possible selection bias which is present if persons entering programs are on average different from comparison group members in their job skills and aptitude.

When an unemployed person participates in an program which does not improve their chance of reemployment there is a *deadweight loss* to society for the spending.[17] When a program participant gains reemployment at the direct expense of an otherwise similar unemployed job seeker then *displacement* has occurred. When an employer, either government or private, receives a subsidy to hire a worker who would have otherwise been hired anyway then *substitution* of program financing for other intended spending has occurred.[18] It is important to consider displacement and substitution effects when doing social benefit-cost assessments of public programs.

Rehabilitating performance monitoring by adopting a system for the ongoing review of short and medium term net/incremental program effects, is an exciting innovation. It offers the prospect of an elegant alternative to the cumbersome system where by performance targets are relaxed by increased service to specially designated hard to reemploy groups.[19] However, new incentive questions are raised by the net/incremental approach. Also, questions about harmonizing the management decision cycle with medium-term outcome measures should be confronted. This includes consideration of budget allocation models based on performance. The Upjohn Institute team will work with all the interested parties in HRDC including representatives from SE&M, EDD, HRIB to review the previous benchmarking work on net/incremental medium-term measures, and to identify appropriate and practical ways to incorporate net/incremental outcome monitoring into a modern performance driven management system for HRDC. To ensure usefulness of the proposed system, we also plan to consult representatives from NHQ, Regional Head Quarters (RHQs), and HRCCs.

---

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

[17] If a program manager practices creaming in selecting participants for programs, then a deadweight loss results.

[18] 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.

[19] Such adjustment methodologies are discussed by O'Leary (1995, 1996).

E.2 Strategies for Drawing Comparison Samples

Key to developing an on going net impact evaluation system for EBSMs in Canada is the issue of how to draw comparison groups. The better job of drawing appropriate comparison groups, the less statistical adjustment will be required to produce reliable net impact estimates of program effects. Issues of relevance include defining the sample frame. This will require a thorough understanding of available data systems and their possible automated interrelationships.

E.3 Issues of Data Completeness and Accuracy

As part of this project it will be crucial to sample data sets from all the systems contemplated for use in the on going net impact evaluation system. The data must be checked for completeness and accuracy. Alternative data sources should be identified where irreparable problems are found. One alternative is follow-up surveys. This is possible, but not desireable. The hope is to define a system around short and medium term measures of appropriate program outcomes based entirely on administrative data which is used in employment offices nationwide.

E.4 Practical Implementation

All design considerations will be done with an eye toward practical implementation of the systems developed. This means partnering with national, regional, and local employment policy practitioners at every stage of the project.

E.5 Uses for Management and Policy Formulation

There is much experience with use of gross outcome monitoring data for program management and planning. The incentive structures and pitfalls are reasonably well understood. Similarly, net impact assessment has been used for policy formulation with great success.

Heckman (1996, p. 2) has noted that: “Performance goals are usually set in terms of levels attained by participants at some point in time. Yet a more meaningful measure of program performance is value-added or gain, which is much harder to measure, but which is the true output of a program.”

It is not known what problems may arise if performance management systems are based on net impact estimates. How will a program manager be expected to use such information to improve the social value of services provided? Furthermore, what are the different uses of short term and medium-term performance measures for program management. In addition to solving technical details of data organization and computation, this project must take great care to consider the important issues of how regular on going net impact estimates will be used on a practical basis.

References

Auer, Peter. 1996. "Monitoring of labour market policy in EU Member states", in *Policies: inforMISEP* (Berlin), No. 53 (Spring), pp. 24-31.

Barnow, Burt. 1992. "The Effects of Performance Standards on State and Local Programs," in *Evaluating Welfare and Training Programs*, Charles Manski and Irwin Garfinkel, eds., Cambridge, Massachusetts: Harvard.

Birch, David L. 1987. *Job creation in America: How our smallest companies put the most people to work*. New York, The Free Press.

Blaylock, Ann B, ed. 1990. *Evaluating Social Programs at the State and Local Level: The JTPA Evaluation Design Project.* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Boeri, Tito. 1997. "Learning from transition economies: Assessing labour market policies across central and eastern Europe," in *Journal of Comparative Economics* (Orlando, FL), Vol. 25, No. 3 (Dec.), pp. 366-384.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences*. Second edition. New York, Academic Press.

Decker, Paul T., and Christopher J. O'Leary. 1995. "Evaluating pooled evidence from the reemployment bonus experiments," in *Journal of Human Resources* (Madison, WI), Vol. 30, No. 3 (Summer), pp. 534-550.

Dickinson, Katherine P., Richard W. West, Deborah J. Kogan, David A. Drury, Marlene S. Franks, Laura Schlichtmann, and Mary Vencil (1988), *Evaluation of the Effects of JTPA Performance Standards on Clients, Services, and Costs*, Washington, DC: National Commission for Employment Policy.

Elias, Peter, and Keith Whitfield. 1987. *The economic impact of the enterprise allowance scheme: Theory and measurement of displacement effects*. Report to the Department of Employment. Coventry, Institute for Employment Research, University of Warwick, Oct.

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.

Fraker, Thomas, and Rebecca Maynard. 1987. "The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs," *Journal of Human Resources*, Volume 22, Number 2 (Spring): 207-227.

Garber, Alan M., and Charles E. Phelps. 1992. "Economic Foundations of Cost-Effectiveness Analysis," *NBER Working Paper Series*, Working Paper No. 4164 (September).

Green, Karen. and Steve Aaronson. 1992. *Performance Measurement in Federal Job, Training, and Education Programs*, Job Training 2000 Performance Standards Subgroup, U.S. Department of Labour, Washington, DC.

Godfrey, Martin, György Lázár, and Christopher O'Leary. 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, Oct.

Góra, Marek, and Christoph M. Schmidt. 1998. "Long-term unemployment, unemployment benefits and social assistance: The Polish experience," in *Empirical Economics* (Berlin), Vol. 23, Nos. ½, pp. 55-86.

Heckman, James J. 1996. *Evaluation Employment and Training Programs: The Case of JTPA*, draft manuscript. Kalamazoo, MI: W.E. Upjohn Institute.

———. 1976. "The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models," in *Annals of Economic and Social Measurement* (Bristol), Vol. 5, No. 4, pp. 475-492.

Heckman, James J., and Jeffrey Smith. 1995. "The Performance of Performance Standards: The Effects of JTPA Performance Standards on Efficiency, Equity, and Participant Outcomes," under revision for *American Economic Review.*

Heinrich, Carolyn J. 1995. "Case Study of the Role of Performance Standards in JTPA Program Administration and Service Delivery at the Local Level," Department of Economics, University of Chicago (mimeo).

Johnson, George E., and James D. Tomola. 1977. "The fiscal substitution effect of alternative approaches to public service employment policy," in *Journal of Human Resources* (Madison, WI), Vol. 12, No. 1 (Winter), pp. 1-26.

Johnson, Terry R. and Ernst W. Stromsdorfer. 1990. "Evaluating Net Program Impact," in *Evaluating Social Programs at the State and Local Level: The JTPA Evaluation Design Project.* Ann B. Blaylock, ed. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Kiefer, Nicholas M. 1988. "Economic Duration Data and Hazard Functions," *Journal of Economic Literature*, 26 (June): 646-79.

Maddala, G.S. 1983. *Limited-dependent and qualitative variables in econometrics,* Cambridge: Cambridge University Press.

Meager, Nigel (with Ceri Evans). 1998. *The evaluation of active labour market measures for the long-term unemployed*. Employment and Training Papers No. 16. Geneva, International Labour Office.

Micklewright, John, and Gyula Nagy. 1998. "Unemployment assistance in Hungary," in *Empirical Economics* (Berlin), Vol. 23, Nos. ½, pp. 155-175.

Mohr, Lawrence B. 1992. *Impact Analysis for Program Evaluation*, London: Sage.

Office of Strategic Planning and Policy Development. 1989. *Guide for Setting JTPA Title II-A and Title III (EDWAA) Performance Standards for PY 89*, Washington, DC: U.S. Department of Labour, Employment and Training Administration.

O'Leary, Christopher J. 1997. "A net impact analysis of active labour programs in Hungary," in *Economics of Transition* (Oxford), Vol. 5, No. 2 (Oct.), pp. 453-484 .

———. 1996. *Methods for performance-based management of active labour programs in Hungary: An adjustment methodology for performance indicators and a proposal for budget allocation*. Kalamazoo, MI, W.E. Upjohn Institute for Employment Research (November).

———. 1996. *Planning Guidelines for Active Labor Programs in Hungary*, a report to the Hungarian Ministry of Labor, Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research (October).

———. 1995. "Performance indicators: A management tool for active labour programmes in Hungary and Poland," in *International Labour Review* (Geneva), Vol. 134, No. 6, pp. 729-751.

———. 1994. *A System for Evaluating Labour Market Programs in Hungary*, A report to the Hungarian Ministry of Labour, W.E. Upjohn Institute Technical Report No. 94-005, Kalamazoo (January).

———. 1993. *Innovation in Planning Hungarian Labour Market Programs*, Labour Market Analysis and Employment Policies Working Paper No. 61, International Labour Office, Geneva (December).

———. 1990. *Evaluation Criteria and Planning Guidelines for Employment Fund Programs in the Republic of Hungary*, Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research (August).

———. 1998a. *Evaluating the effectiveness of active labour programs in HUNGARY*. Technical Report Number 98-13. Kalamazoo, MI, W. E. Upjohn Institute for Employment Research..

———. 1998b. *Evaluating the effectiveness of active labour programs in POLAND*. Technical Report Number 98-12. Kalamazoo, MI, W.E. Upjohn Institute for Employment Research.

O'Leary, Christopher J., Piotr Kolodziejczyk and Gyorgy Lazar. 1998. "The Net Impact of Active Labour Programmes in Hungary and Poland," *International Labour Review,* Vol. 137, No. 3.

O'Leary, Christopher J., and Andrew S. Targowski. 1993. *MIS Guidelines for Employment Law Programs in Poland,* Kalamazoo: W.E. Upjohn Institute for Employment Research.

Osborne, David and Ted Gaebler. 1992. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*, Reading, Massachusetts: Addison-Wesley.

OECD. 1994. *Performance management in government: Performance measurement and results-oriented management*. Public Management Occasional Paper 1994 No. 3. Paris.

Puhani, Patrick A. 1996. *Poland on the dole: Unemployment benefits, training and long term unemployment during transition*. Discussion Paper No. 96-30. Mannheim, Centre for European Economic Research.

Puhani, Patrick A., and Victor Steiner. 1996. *Public works for Poland? Active labor market policies during transition*. Discussion Paper No. 96-01. Mannheim, Centre for European Economic Research.

Ryan, Kenneth and Ronnie Kauder. 1990. *Measuring Up: Planning and Managing with Performance Standards PY 1990*, Washington, DC: National Alliance of Business.

Simpson, Carl. 1990. "Evaluating Gross Program Outcomes," in *Evaluating Social Programs at the State and Local Level: The JTPA Evaluation Design Project.* Ann B. Blaylock, ed. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Singer, Burton. 1986. "Self-Selection and Performance Based Ratings: A Case Study in Program Evaluation," in *Drawing Inferences from Self-Selected Samples*, Howard Wainer, editor, New York: Springer-Verlag.

U.S. Congress. 1998. *Federal Register,* Vol. 63, No. 113 (Friday, June 12).

U.S. General Accounting Office, Health, Education and Human Services Division. 1998. *Results Act: Observations on Labor's Fiscal Year 1999 Performance Plan*. Report GAO/HEHS-98-175R. Washington, DC (June).

Wandner, Stephen (ed.). 1992. *Self employment programs for unemployed workers*. Unemployment Insurance Occasional Paper 92-2. Washington, DC, US Department of Labor, Employment and Training Administration.

West, Richard W. 1992. *Development of Adjustment Models for PY 92 JTPA Performance Standards for Titles II-A and III*, Menlo Park, CA: Social Policy Research Associates (June).