

**An Overview of the Summative Evaluations of
Employment Benefits and Support Measures
Under the
Labour Market Development Agreements**

Walter Nicholson

DRAFT 5

August, 2008

The *Employment Insurance Act* of 1996 revised unemployment insurance policy in Canada to include two parts. Part I of the Act provides Employment Insurance (EI) benefits to workers who have lost their jobs. Part II provides employment benefits and support measures (EBSM) to help current and former EI clients return to work. Although Part I benefits are administered by the federal government, the 1996 Act gave provincial and territorial governments significant roles in operating the Part II programs. Under the Labour Market Development Agreements (LMDA) mandated by the Act such responsibility can be either fully transferred to the provinces and territories, or the programs can be “co-managed” and operate under shared responsibilities.

The 1996 Act also required that programs operated under the LMDAs be subject to formal evaluations over the ensuing decade. Many of these evaluations have now been completed. Typically the evaluations have been undertaken in two stages. The first, “formative” evaluations looked at how the LMDAs were implemented and examined in detail some of the programs being operated under them. A second round of “summative” evaluations followed. These focused on measuring the impacts that the programs had. The purpose of this overview report is to summarize and critique these summative evaluations, primarily with respect to their attempts to estimate the quantitative impacts of the EBSM programs. This report focuses explicitly on the six quantitative evaluations that had been completed by year-end 2007 – those for Alberta, British Columbia, New Brunswick, Newfoundland and Labrador, Ontario, and Quebec. As new evaluations are completed, the tables from this report will be updated to reflect the additional estimates.

This report is divided into seven major sections. Section I provides some general background on active labour market programs and describes how the EBSM summative evaluations fit into the general evaluation literature for those programs. Issues in methodology are discussed in Section II with specific attention to the advantages and shortcomings for the designs developed for the summative evaluations. A similar set of questions about the data resources available in the evaluations are taken up in Section III. Section IV, the longest in the report, summarizes the quantitative impact results from those summative evaluations that have been completed to date. The goal of this summary is to highlight the findings of the evaluations in a compact format, to offer some conjectures about why the results turned out the way they did, and to place the estimated results into a comparative perspective. Section V then provides a concise summary of the qualitative research in the evaluations in an effort to provide some context to the quantitative results. Issues surrounding the cost-benefit and cost-effectiveness analyses in the evaluation are discussed in Section VI. Finally, Section VII summarizes the lessons learned from the summative evaluations both with respect to the effectiveness of various active labour market policies and with respect to the effective design of ways to evaluate them.

I. Background on Active Labour Market Programs and their Evaluation

Most developed countries sponsor extensive active labour market programs. A recent survey of OECD countries finds that on average these countries spend about 2-3 percent of GDP on labour market programs with about two-thirds of that going toward

“passive” programs (primarily unemployment compensation) and one-third toward active measures (Martin and Grubb, 2001). Spending on all labour market programs has a significant anti-cyclical pattern. As might be expected, passive benefits expand most rapidly during cyclical downturns. The authors point out that despite considerable political rhetoric favoring shifting benefits toward active programs, actual movements in that direction have been very limited – perhaps because of “doubts about the effectiveness of much of this spending.”

The content of active labour market programs is remarkably similar across the OECD countries. In most countries the largest share of spending is devoted to formal training programs, usually in a classroom setting. Such training can be provided directly through government agencies, or indirectly through third parties. In some cases clients are asked to pay a portion of their training costs.

Subsidies to private sector employment constitute a growing component of active programs. Often such subsidies are targeted at disadvantaged workers, are of a limited duration (six months to one year), and are expected to have significant on-the-job training opportunities. Whether such subsidies encourage employment after the subsidy period ends is a focus of much research. The issue of whether subsidized employment causes large displacement effects for non-subsidized workers is also a major concern.

Many countries also encourage direct job creation in the public or non-profit sector. In some cases such jobs are used as a test of labour force attachment for recipients of unemployment compensation. Whether the jobs produce valuable outputs and/or useful labour market experiences for participants have been hotly contested questions.

Assistance in pursuing self-employment opportunities is a less common, though growing component of many nations’ active programs. This approach is based on the presumption that, for a relatively narrow subset of clients, starting businesses may be a direct route to meeting their own needs and, potentially, increasing jobs for other workers. The popularity of the self-employment option has tended to obscure that fact that very little is known about its long term efficacy, however.

Finally, all countries offer a variety of job search assistance activities, either through a formal public employment service or through contracting with third parties. Services may include access to job postings, resources for making employer contacts, assistance in resume preparation or other job-related counseling, or direct job placements. In some cases these activities may be monitored as a way of enforcing availability for work requirements in unemployment compensation laws. Although job search programs usually serve more workers than any other active labour market program, per-client costs are usually relatively low.

Canadian experiences closely mirror those of the typical OECD country. Spending on labour market programs has averaged around two percent of GDP during the past twenty years. Expenditures on active labour market programs have remained relatively constant at about 25 percent of total labour market spending. In all, such active

spending amounts to about 0.5 percent of GDP according to the OECD data. But these data are quite comprehensive and include such elements as special youth programs and administrative cost of the employment service. Spending on the EBSM program alone has been about \$2 billion over the past several years¹ – that is, approximately 0.15 percent of GDP.

Services available under the EBSM programs also resemble the offerings of the other OECD countries generally. Although different provinces may use different names and slightly different definitions for their interventions, they usually include the same five activities common to most other countries². In order to simplify the presentation, throughout this report we will focus on five specific EBSM interventions:

- **Skills Development (SD):** These are the primary training activities offered under the EBSM programs. Typically this training is funded through client vouchers to third party providers and clients pay a negotiated portion of the total cost.
- **Target Wage Subsidies (TWS):** These subsidies are typically targeted to hard to employ workers and may be up to 60 percent of total wages. The subsidies last for up to 52 weeks and can be extended to 78 weeks for workers with disabilities. Earnings from subsidized jobs are insurable under EI and therefore enable workers to renew their EI eligibility.
- **Self Employment (SE):** Under this intervention workers are provided with assistance in setting up their own businesses. They are able to collect their remaining EI entitlements during this process and may in some cases collect additional Part II benefits. Earnings under SE are not insurable under EI and therefore do not provide for renewed EI eligibility.
- **Job Creation Partnerships (JCP):** These are jobs in the public and non-profit sector that are intended to benefit the community. Employees on these jobs continue to receive their EI benefits in lieu of wages and may have these benefits “topped-up” to prevailing wage rates for the specific occupations. Wages earned under JCP activities are not insurable under EI.
- **Employment Assistance Services (EAS):** These are job search assistance services provided to help unemployed workers find employment. The services are often provided through third-party service delivery agreements.

Because these interventions are intended to operate quite differently on a conceptual level, we will generally study them separately (although some evaluations have provided aggregated estimates of the “effect of EBSM”). This disaggregated approach is also the approach taken by the extensive literature on the evaluation of active labour market programs. The vast bulk of this literature has focused on formal training

¹ This figure does not include client contributions for their interventions, nor does it include client opportunity costs.

² Some authors point to a sixth type of active labour market program: youth-oriented activities. In Canada these are operated separately from the EBSM program and hence such programs are not a subject of this report.

programs. An excellent summary is provided by Heckman, LaLonde, and Smith (1999) and we shall make extensive use of the findings from this survey in what follows. In general the authors find that training can provide gains to employment and earnings that are roughly consistent with the returns to formal education. But the results from training evaluations are quite variable depending on the nature of the training, the intended target group, and the methodology used to measure impacts. Because training programs are relatively expensive (especially once opportunity costs are considered), only a few evaluations find that the benefits of these programs exceed their costs.

Heckman, LaLonde, and Smith also provide a more limited discussion of the evidence from job search programs. A more extensive treatment of the U.S. experience is provided in Meyer (1995). This paper reviews a series of (random assignment) job search experiments. Overall, these summaries suggest that job search assistance may have modest beneficial impacts on employment-related outcomes. Because these are low-cost interventions, however, it is sometimes the case that such programs will have benefits that exceed costs.

Evaluations of employment subsidies, self-employment, or public sector job creation are far less common. Martin and Grubb (2001) attempt an overall assessment of such activities, but reach rather ambiguous conclusions. An important benefit of the EBSM evaluations is therefore to shed more light on these less-studied interventions.

II. Methodology of the Summative Evaluations

The Labour Market Development Agreements provide explicitly for a quantitative analysis of the effectiveness of EBSM interventions. Provision of such estimates is a primary goal of the summative evaluations. It was recognized at the outset that obtaining reliable estimates of the effects of on-going programs raises many difficult methodological issues³. In this section we summarize those issues and show how they were addressed in the various summative evaluations.

A. Assessing Impacts: The Need for a Counterfactual

In order to estimate the effect that a program has on those who participate in it, one must specify a “counterfactual” that identifies what would have happened to those individuals had they not participated. A wide variety of approaches have been taken to this issue over the past forty years of program evaluation. In many ways the cleanest approach to the problem uses random assignment. When program participation is determined randomly it may be reasonable to assume that the experiences of those who were randomly denied services provide a good picture of what would have happened to participants had they not entered a program⁴. In the EBSM context, however, the random

³ Many of these issues are described in Nicholson (2001).

⁴ Of course, this simple presumption may not hold in many actual implementations of random assignment. For example, if some individuals who are randomly offered a program choose ultimately not to participate, one must evaluate the impact of being offered a program which may differ substantially from the impact of actually participating in it.

assignment approach to evaluation was not feasible because the denial of service required to implement it would violate legal principles of free access to these programs.

Because random assignment was infeasible in the EBSM context, analysts were required to adopt “non-experimental” methods of evaluation. The need to adopt such methods did not, of course, eliminate the need for a counterfactual, but it did require focusing almost exclusively on designs that use comparison groups for that purpose. Various types of comparison groups have been used for many years in evaluations. The primary issue in using them is whether the experiences of the groups chosen do indeed provide a good measure of what would have happened to program clients had they not participated in the program. In other words, are members of the comparison group sufficiently similar to members of the participant group so that any difference between their experiences can be attributed to the program?

Ultimately the only way to choose a comparison group that seeks to be similar to a group of participants is to use whatever data are available to ensure that the groups are similar on measurable dimensions. Although there are many approaches to doing this (some of which are quite sophisticated), one should not lose track of the basic point that any analyst is constrained by the data that are available. If participants and members of the comparison group differ significantly along dimensions that cannot be measured, the validity of impact results based on comparing the two groups is thrown into doubt.

B. The Choice of Matching as a Methodology

All of the summative evaluations chose some form of statistical matching as a way of selecting a comparison group. Usually this was done in two stages, based on what data were available at each stage. In the first stage, the analysts identified a sample of program participants and collected administrative data on their basic demographic and labour market characteristics. Because these data were also available for otherwise eligible individuals who did not participate in EBSM interventions over a specific period⁵, they could be used to select a comparison group from these non-participants that resembled the participant sample. This was usually done on a one-to-one matching basis. That is each participant was matched to one or more non-participants on the basis of the available data. In part this matching was done as a “hard match” – that is the participant and comparison case had to agree precisely on the variable being measured. The most common variables used in this way were gender, age, and local labor market. Additional matching was accomplished through the use of “propensity scores”. That is, the available data were used to estimate a probability of participating equation and that equation was used to develop participation probabilities for individuals in both the participant and

⁵ Specification of a “non-participant” sample raises several definitional issues in the context of an on-going program such as EBSM. The two most important of these are: (1) How is past participation to be handled in selecting the participant and comparison samples? (2) How is the possibility of future participation to be handled for the two samples? The summative evaluations appear to have adopted a variety of answers to these questions. In principle, any criterion that is applied to the participant group with regard to past or future participation should also be applied to the comparison group. For example, if participants with past interventions are excluded, comparison cases should be also. Future participation can be treated as an outcome of interest, but this outcome was seldom examined.

comparison groups. Each participant was then matched to his or her “nearest neighbor” in terms of propensity score.

Few of the summative reports provide precise details on the results of this first stage matching procedure, perhaps because such details would not be of interest to general readers. But much of the validity of the subsequent analysis hinges on the quality of this initial match. There is evidence, for example, of the possibility that some matching with respect to labour market characteristics was imperfect⁶. Of course, of necessity the first stage matching was rather crude, primarily because only a very limited number of variables were available. Whether potential problems were subsequently ameliorated through the second stage estimation procedures that were employed remains an open question.

First stage matching provided the sample frame for surveys that were conducted in all of the evaluations. These surveys provided much more detail on demographics, education, and labor market histories than was available from administrative data. Most surveys were conducted between 18 and 24 months after EBSM participation ended, so, in principle, they could collect labour market and other information over three general time periods: (1) Pre-program; (2) In-program; and (3) Post-program. The pre-program data collected in this way were used as explanatory variables in the second stage estimation procedures whereas the other data were used to measure within program outcomes (that is, opportunity costs) and post-program impacts. The evaluations varied with respect to the extent to which they relied on the survey data for most of their analysis or relied primarily on administrative data⁷. In general, evaluations conducted later tended to make more use of administrative data, especially in cases where CRA data became available. Evaluations using CRA data generally had longer follow-up periods than did those using survey data and there is some indication that key variables such as earnings were measured more accurately in the administrative data. Because the CRA data were only available on an annual basis, however, their use did introduce some fuzziness into defining outcomes during pre- and post-program periods.

The evaluations adopted a variety of second stage estimation procedures to arrive at their final impact estimates. Although these may seem quite different, it is important to recognize their overall similarity. All of the second stage methods used whatever additional data the analysts had available in attempting to hold constant all measurable differences between individuals in the participant and comparison groups. Though these

⁶ For example the evaluations for Alberta found that participants had significantly greater EI collections in the period immediately prior to program entry than did matched non-participants. Also in some cases participants appear to have somewhat less base period employment than do individuals in the comparison groups. This issue is discussed further in connection with problems in dealing with the “Ashenfelter Dip” below.

⁷ The survey data tended to suffer from two drawbacks. First, as is common for all surveys, survey response rates tended to be around 40 percent, so analysts were forced to make some adjustments for sample attrition (see the discussion of weighting below). Second, much of the retrospective labour force data may have suffered from response errors. That was especially true for the pre-program data, some of which related to events more than four years before the survey. Such data problems are discussed in Section III

procedures may indeed help to control spurious correlations between measurable differences and observed outcomes, none of them can control for remaining unmeasured differences between the groups. Hence, none of them is fully immune to criticisms that the impact estimates may suffer from selectivity biases – a problem inherent in all non-experimental evaluations.

Four specific second stage estimation procedures were used in the evaluations:

- Ordinary least squares estimation⁸ (British Columbia, Labrador and Newfoundland, Ontario);
- Nearest neighbor propensity score matching (Quebec); and
- Kernel matching⁹ (Alberta, British Columbia, New Brunswick, Ontario, Saskatchewan).
- Propensity score weighting (Nova Scotia, Prince Edward Island)

In principle these estimation procedures could be applied either to estimate post-program differences in mean outcomes between the participant and comparison groups (“cross section” estimates) or to estimate differences in the pre-post program change in outcomes (“difference-in-difference” estimates). In the end, virtually all of the evaluations focused on the difference-in-difference (DID) estimates although in some cases cross-section values were reported as well (Alberta, Nova Scotia, Prince Edward Island, and Quebec). Most of the reports found that the cross-section and DID estimates were “similar”. However, that was not always the case, in part because of the difficulty in obtaining good baseline data from the surveys needed to calculate the DID outcome variables (see below).

C. Temporal alignment of the participant and comparison samples

All of the evaluations defined their participant samples on the basis of the end dates of their primary interventions¹⁰. Because many individuals were missing such dates, it was customary to assume that if no activity occurred over a six month period, an intervention must have ended at its last recorded activity. Although this assumption incorporated an admitted uncertainty into the definition of the intervention period and, more importantly, into the definition of the “post-program” period, the choice was a reasonable one to make.

⁸ This technique was supplemented with an instrumental variable approach in British Columbia using a distance measure as an instrument. This approach was unsuccessful, however and results derived from it are not reported. The Quebec report also mentions use of an instrumental variable procedure, but details are not provided.

⁹ Kernel matching constructs a hypothetical comparison case for each participant by weighting the members of the comparison group on the basis of the distance of their propensity scores from that of the participant. British Columbia provided both OLS and Kernel matching estimates.

¹⁰ Usually this was defined as the longest intervention within the sample period. Claimants with multiple interventions were treated as having an “action plan” that included this full package, but most analysis was done only on the principal intervention. Possible treatment interactions for those with multiple interventions were not examined extensively, though theory suggests that some intervention combinations (i.e. TWS followed by SD) might be especially productive.

Once the participant sample was selected, data on intervention start dates were collected and these were used to define the “pre-program” period. The approaches taken to define this period varied. For active claimants this period was usually defined as ending at the EI start date for the claim that preceded entry into the EBSM intervention whereas for former claimants the pre-program period was usually defined as ending at the intervention start date.

Because members of the comparison group, by definition, did not have start and end dates for interventions, it was necessary to impute hypothetical dates in order to define pre- and post-program periods. Before looking at these procedures in detail, it is important to understand the reasons to be concerned about them. In the case of the definition of the pre-program period, the major concern is controlling for what has come to be known as the “Ashenfelter earnings dip” (see Ashenfelter, 1978; Heckman and Smith, 1999). The fact that earnings of participants tend to show a pronounced decline shortly before entering an active labour market program has been observed in practically all evaluations. Presence of the “dip” is clear evidence that individuals carefully consider their own circumstances in deciding when to enter a program. This self-selection poses a challenge to those who would evaluate such programs using non-experimental methods. In its most simple incarnation, failure to eliminate the influence of the earnings dip phenomenon in the participant sample would make it appear that program participants had outsized earnings gains (and related large declines in EI use) in DID estimates. To control for this one must be very careful to model accurately the labour force dynamics of members of the comparison group so as to mirror these experiences of participants.

For active participants, use of the EI start date (rather than the intervention start date) was intended in part to address the dip phenomenon¹¹. Choosing a comparison sample with similar EI start dates then provided a crude approximation to the dip phenomenon. Further control of the problem can be obtained by matching on the length of the EI spell prior to the intervention start and this procedure was used in British Columbia, Labrador and Newfoundland, and Ontario. In general, then, the extent to which the dip phenomenon may have affected outcome estimates for the active sample must be evaluated on a case-by-case basis.

Alignment of the participant and comparison samples was more difficult for former claimants than for active ones. In this case it was thought that using EI start dates would generally be too distant for proper alignment, though this might have worked for fairly recent former clients. A solution followed in the Newfoundland, Nova Scotia, Prince Edward Island, and Ontario evaluations was to limit the comparison group to individuals who had participated in EAS only¹². In these cases then the EAS start date for the comparison sample was aligned with the start date for the major intervention of

¹¹ EI data were used to control for the earnings dip phenomenon because direct earnings measures are not available in sufficient temporal detail from administrative sources (i.e., the CRA data only provide annual earnings data).

¹² The Nova Scotia and Prince Edward Island evaluations did not explicitly differentiate between active and former claimants and therefore used the “EAS only” group as a comparison for both types of participants. The authors believe this choice of a comparison group may have helped to control for the significant amount of seasonal employment in their samples.

the participant. Impacts measured in this way should be interpreted as being incremental – that is, in excess of whatever impact the EAS intervention had. Other evaluations adopted more-or-less *ad hoc* procedures for aligning the participant and comparison samples for former claimants. A common requirement was that comparison cases with similar propensity scores to a participant were also required to be “unemployed” (actually “not employed”) at that participant’s intervention start date. Usually a similar screen was not used for the participant, however, because that would have eliminated a number of participants from the evaluations.

Issues in the proper definition of the “post-program period” have not been explored in much detail in the evaluations. Two facts seem clear. First, if the definition of the post-program period allows for some continuing program participation by members of the participant sample, outcome measures based on employment or earnings will be understated. Second, similar biases can be incorporated into the data for the comparison group depending on precisely how a hypothetical “program end date” is imputed. Both of these problems may be ameliorated by using a post-program period that is arbitrarily defined to be well after program participation has ended for both comparison and participant samples. This procedure was followed in some of the evaluations, though limitations in the length of the follow-up period (especially with respect to the survey data) constrained the options available in many cases.

III. Data resources in the evaluations

All of the evaluations collected data from a variety of sources. This section briefly discusses four such sources¹³:

- Administrative data on EI;
- Administrative data on earnings;
- Survey data;
- Qualitative data.

We show why the analysts tended to rely increasingly on the administrative data in the later evaluations.

A. Administrative data on EI

The Status Vector file provided the primary source for data on EI collections and related HRDC files provided the primary ways of identifying participants in interventions. These data generally proved to be some of the most reliable in all of the evaluations. Although in some cases there were problems in identifying the population of all participants with EBSM end dates during a defined period (in part because the data on end dates were incomplete), eventually most of the samples obtained were reasonably representative and complete.

The Status Vector data on EI benefit collections proved to be invaluable both as a source for data with which to match participants and comparison group members and as a

¹³ Sources of data on program costs are discussed separately in Section VI below.

source for assessing impacts on EI. It is the detail on the timing of EI collections that makes these data so valuable. Without such detail it is likely that the matching procedures used in the evaluations would have been impossible.

However, mistakes on both conceptual and practical levels may have been made in using the Status Vector data. On a conceptual level, because EI benefits are paid for a variety of circumstances, the analyst must select those EI components deemed relevant to the question being asked. For example, although it seems clear that reductions in collections of regular EI benefits can be regarded as a beneficial outcome of EBSM interventions¹⁴, the case is not so clear for impacts on maternity, parental or other special benefits¹⁵. In his validation report on the Alberta evaluation Szabo (2007) shows that estimated EI impacts can be quite different depending on precisely which benefits are included in the totals. But most evaluations simply used total benefits.

On a practical level, some care must be taken in using the Status Vector files to ensure that the correct series are being used. For example, it seems to have been common practice in the evaluations to use total weeks of benefits collected as reported in the “header” files of the Status Vector data. Szabo (2007) also shows that this can in some cases give misleading measures for the impact of EBSM interventions on EI collections. This is especially true when interest centers on EI collections during a specified time period (i.e. during the first post-program year). In order to get such measures precisely correct, one must use the detailed period-by-period data contained in the “trailer” records.

In general it appears that such problems may have been relatively inconsequential in terms of the results. But future evaluations should take advantage of the lessons learned here about the use of the Status Vector Files.

B. Administrative data on earnings

Tax data on earnings from the Canada Revenue Agency (CRA) in principle would have provided a good source of both matching and outcome data for the evaluations. Unfortunately, concerns about confidentiality generally prevented the use of these data for most of the evaluations reviewed here. The New Brunswick, Nova Scotia, and Prince Edward Island evaluations did make extensive use of CRA data, and it seems likely that the later evaluations will do the same. Although use of these data does pose some problems arising from their temporal aggregation (the data are available only on an annual basis), it seems that they may provide more reliable data than surveys because non-response and reporting errors may be much smaller. Use of the CRA data to verify some of the results from the earlier evaluations might also be considered (see Section VII).

C. Survey data

¹⁴ Even for collection of regular benefits it is important to make a distinction between reductions caused by increased employment and reductions induced by lack of eligibility (see Section IV).

¹⁵ In total such special benefits represent about one-third of all EI benefits.

All of the evaluations conducted surveys to gather information on such topics as education and skills, employment patterns, and perceived experiences with EBSM interventions. Sample sizes for these surveys varied across the evaluations, but typically included about 2,000 participants and a like number of comparison cases. Usually a stratified sample design was used for the surveys. The purpose of such stratification was to ensure adequate representation of the numerically smaller interventions and, in some cases, to obtain geographic diversity¹⁶. Because of this complex sample design, most analysts used weighting procedures for reported results so that characteristics of the sample would more accurately reflect the characteristics of the underlying populations¹⁷.

Evaluations that relied on survey data for most employment-related measures tended to encounter some data problems. Two issues were most salient. First, as for many surveys, the response rates for the evaluation surveys were low – typically well below 40 percent. Reasons for such low rates include the mobility of the population being surveyed, poor quality of contact information, and an increasing unwillingness of individuals to participate in any surveys. The concern, of course, is that these effects may result in a sample that is both unrepresentative of the underlying population and yields misleading estimates of program impacts.

It is difficult to determine precisely how the evaluations dealt with the non-response issue on the basis of the published reports. Although it was possible to do a fairly extensive examination of the issue using the administrative data that were employed to select the survey sample, details of such an analysis are reported only in the British Columbia evaluation (*Annex G*). There the authors find that response rates were higher for females, for older workers, and for those with EI benefit receipt in the thirteen weeks before the start of the intervention. Response rates were lowest for those who had SD as their primary intervention or who were EAS-only clients. These findings provide some evidence that non-response was not random, but followed clear and relatively predictable patterns. Similar analyses in most of the evaluations led the authors to adopt weighting schemes for much of their impact analyses. Precise choices for these weights had the potential to affect the impact estimates significantly.

A second issue arising from the use of survey data is the possibility that some of the retrospective data on employment and earnings may suffer from recall biases. The possibility for such errors was clearly recognized in many of the design documents for the EBSM evaluations (Nicholson, 2001). The fear was that, because the surveys needed to be administered perhaps two years or more after the interventions ended, the data on pre-program employment and earnings could be four or more years old. The main consequence of such reporting errors would be to add variability to difference-in-

¹⁶ Whether the resulting designs were efficient in the sense of measuring key outcomes with minimum variance is an open question since most of the evaluations did not explore the size of the “design effect” introduced by their stratifications. There are some suggestions that such designs may have induced some inefficiency. For example, British Columbia developed a complicated sample design allocation which, because of the constraints involved, resulted in requiring large portions of the skills development sample to be concentrated outside of the Vancouver area.

¹⁷ Weighting was also used to adjust for sample non-response. Issues surrounding use of weights in the evaluations are discussed in Appendix 1.

difference measures of the key employment related outcomes. Errors in the pre-program variables could also have an influence on the second stage matching procedures used.

Because the early evaluations did not have access to CRA data, there was no independent way to assess the reliability of the survey data. An extensive examination of the data from New Brunswick (Waslander, 2007) shows that there are some discrepancies between employment and earnings reported on the survey and what can reasonably be inferred from EI and CRA administrative data. After a detailed set of edits the author shows that only about two-thirds of respondents' reports on employment and earnings during post-program years are in accord with the administrative data. Only about half of the kind of before-and-after data that might be used in the development of difference-in-difference estimates met a similar set of tests. Unfortunately this analysis did not focus explicitly on the type of pre-program data that is thought to be most vulnerable to response errors. But the findings do suggest some care in relying exclusively on survey data for impact analyses.

Non-response and reporting errors are common to all survey data. However, this has not prevented many analysts from relying on them. For example, many of the most influential evaluations of active labour market programs in the United States (such as the Supported Work or the JTPA evaluation) have been based almost exclusively on survey data, with relatively little concern about the quality of the data. Hence, reliance on survey data in the EBSM evaluations is not anomalous. It might, however, be useful to undertake a thorough investigation of the actual survey instruments used in the evaluations and of the procedures employed to increase response and recall accuracy to determine whether these were of an acceptable quality. Some suggestions for this research are developed in Section VII.

D. Qualitative data

All of the evaluations collected a considerable amount of qualitative data from three general sources:

- Extensive reviews of documents about the EBSM programs
- Key informant interviews; and
- Focus groups of participants, employers, and community leaders.

These data were used for two purposes. First, they were intended to provide a more general context to the interventions being studied in the quantitative analysis. That is, they were used to show how the interventions actually worked in practice and to shed some light on why the results turned out in the ways they did. In Section V we discuss these insights in the context of the actual quantitative impact estimates obtained.

A second use for the qualitative data was to address topics that could not readily be handled in the quantitative framework. The two principal such topics were: (1) Effects of EBSMs on employers; and (2) Community effects of EBSMs. Conducting these

studies posed significant methodological challenges in the evaluations. A discussion of the issues is provided in Appendix 3.

IV. Quantitative results of the evaluations

Viewed as a whole, the EBSM evaluations represent one of the largest studies of active labour market programs ever undertaken. The evaluations generated literally hundreds of estimates for a wide variety of interventions, outcome measures, and client groups. It would be impossible to summarize all of these estimates here. Rather, the goal of this section is to focus on a few representative estimates and to seek to place these into a more general overall context. Potential ambiguities about the estimates will also be discussed in some detail. Policy lessons that can be drawn from the estimates will be discussed in Section VII.

A. Economic context of the evaluations

Before looking at the results from the EBSM interventions it may be helpful to summarize some background data on the contexts within which the interventions operated. First, consider local labour markets. Table 1 looks at unemployment rates in the provinces where the EBSM evaluations occurred. The data are organized around each evaluation's schedule. That is, they show unemployment rates that prevailed during the "pre-program" year, the "intervention" year, and during two "post-program years. Although this chronology does not line up perfectly with the actual dates used in the evaluations, the correspondence is close enough to give a general impression of labour market trends.

Several facts are immediately apparent in Table 1:

- The labour market in Alberta was strong throughout the period;
- Labour markets in British Columbia and Quebec were relatively weak throughout the period;
- The labour market in Ontario became weaker over the evaluation period; and
- The labour markets in the Maritime Provinces (New Brunswick, Newfoundland, Nova Scotia, and Prince Edward Island) were very weak throughout the evaluation period.

These differences raise the question of whether the labour market environment alone might have influenced the evaluation results. There is some debate in the literature about whether active labour market programs are more effective during periods of strong labour markets or weak ones. In strong markets program participants will more readily find jobs once their interventions end. But strong markets also benefit members of the comparison group so that the opportunity costs of being "locked in" to a program for a time will be greater.

Perhaps the best evidence on this question comes from a recent study of detailed administrative data from the German labour market over the period 1980-2003 (Lechner and Wunsch, 2006). These authors find long-run employment gains of about 10 percent from participation in active labour market programs with roughly equivalent gains in earnings. They also report a significant negative correlation of estimated employment and earnings impacts with the state of the labour market. Overall gains are between 0.7 percent and 1.8 percent greater for each one percentage point increase in the national unemployment rate. Similar effects are also reported for differences across regional labour markets in Germany. Hence, it seems plausible that the EBSM programs might prove to be more effective in the provinces with higher unemployment rates -- British Columbia, Quebec, and possibly in the Maritimes (though the unemployment rates there are considerably higher than in the German study).

B. Characteristics of program participants

Estimated impacts of EBSM interventions may also have been affected by the characteristics of program participants. Table 2 provides a brief summary of these characteristics by province¹⁸. Three features of these data might be explicitly highlighted. First, the fraction of participants who were female varied significantly across the evaluations, ranging from a low of 36 percent in Nova Scotia to a high of 53 percent in Alberta. Second, participants were considerably younger in the Maritimes than in the other provinces. Finally, participants in all provinces were relatively well-educated. Only in New Brunswick did fewer than 30 percent of participants have some education beyond high school¹⁹.

Ideally one might like to see the figures in Table 2 disaggregated both by claimant status and by specific intervention. Such a breakdown would highlight the precise ways in which characteristics of participants may have affected the results. Because of limitations posed by available sample sizes, the evaluation reports do not generally provide such details. In some cases, however, the reports do point out major differences on a few dimensions. For example, it seems clear that women constitute a higher fraction of the caseload among former than among active claimants. Similarly, women were a bit more prevalent in the TWS and JCP interventions and somewhat less in SD and SE. Participants in the self-employment intervention also had considerably higher levels of formal education than did participants in the other interventions.

C. Rationale for selecting results

Four principles underlie the selection of results to be reported here:

¹⁸ Because the evaluation reports did not use the same break points in reporting distributions, it was necessary to approximate some of the data in Table 2. Data for Quebec were not currently available.

¹⁹ Although these levels of educational attainment do not seem significantly different from those of the overall labour forces in the provinces, they do suggest that the EBSM program is not explicitly targeted on disadvantaged workers as is the case for many active labour market initiatives in other countries. The high levels of educational attainment have also caused some analysts to worry that the EBSM interventions may engage in “creaming” in enrolling participants. The evaluations that have considered this issue, however, have generally discounted its importance in affecting the overall results.

1. **Only results estimated separately by intervention will be reported.** Because the interventions studied in the evaluations have very different conceptual bases, it seems likely that each is characterized by a unique set of structural determinants for outcomes. Although some of the evaluations reported an “overall” effect for the EBSM package of interventions, these will not be discussed.
2. **Outcomes will be presented separately for active and former claimants²⁰.** There are two reasons for this categorization. First, active claimants may be better attached to the labour force than are former claimants. Hence they may respond quite differently to the various interventions. Second, because of these different characteristics, most researchers found it more difficult to develop a comparison group for the former claimants than for the active claimants. There is then the possibility that the estimates for former claimants may be less accurate.
3. **Only three specific outcomes will be examined.** Although most of the evaluations measured six or more primary quantitative outcomes, here we will only summarize three of these: (1) Employment – measured in annualized hours worked; (2) Earnings – measured in dollars per year; and (3) EI collections – measured in weeks per year²¹. These three outcome measures are commonly used in active labour market evaluations and choosing to focus on them here is in part to facilitate comparisons with such other studies²².
4. **Difference-in-difference estimates will be emphasized.** Most of the evaluations reported estimated impacts in two ways: (1) Simple cross section differences between the participant and comparison groups; and (2) Difference-in-difference estimated for the two groups. Usually the difference-in-difference (DID) results were emphasized in the final reports and we will follow that practice here. In principle such estimates do have an advantage over the cross-section figures because time invariant differences between the participant and comparison groups are controlled for in such estimates. The advantages of DID estimates might be less than believed in the evaluations, however, because of problems in measuring

²⁰ The Nova Scotia and Prince Edward Island evaluations did not differentiate by claimant status. Results from these evaluations will be discussed under “Active Claimants”. The Newfoundland evaluation disaggregated the group of former claimants depending on social assistance status. Those with prior social assistance collections were termed “common clients”. The results reported here for Newfoundland focus only on those former claimants who were not common clients.

²¹ Because not all of the evaluations measured outcomes in these ways, some re-calibration of results was necessary. Other adjustments were made for post program time periods actually covered so as to put the annual figures on a consistent basis.

²² Three commonly reported outcomes that will not be discussed here are: (1) Dollars of EI benefits; (2) Dollars of Social Assistance benefits; and (3) Weeks on Social Assistance. The first of these is not reported because it is largely duplicative of the EI weeks figure. Social Assistance related outcomes are very difficult to interpret in the evaluations, in part because estimating these posed significant methodological challenges – especially in finding an adequate comparison group. Some of these issues are discussed briefly in Appendix 2.

the pre-program levels of the outcomes of interest²³. Hence we will occasionally mention some situations in which the cross section results help to clarify matters²⁴.

D. Results for Skills Development (SD)

Skills development is the most expensive intervention in the EBSM package. In 2005/06 spending on SD (\$919M) amounted to nearly 50 percent of total spending on all EBSM services (*Monitoring and Assessment Report, 2007*). These benefits were provided to about 133,000 Canadians²⁵, so the implied cost per client was approximately \$6,900 per person. Clearly this program involves significant levels of human resources investment, so it might be expected that there would be considerable interest in the results for these clients.

In theory the SD intervention is the most like a pure investment in human capital of any of the EBSM programs. The participant spends significant time learning a set of skills (usually in a classroom setting) and, in the majority of cases, earns some type of certificate or diploma. This intervention is therefore similar to formal education. Since extensive studies of formal education suggest that each additional year yields approximately a ten percent increase in annual earnings, a rough guess might be that SD participants (who typically spend about six months in their interventions) might have increases in earnings on the order of about 5 percent or so. But actual experiences with such training programs have generally not been this positive. There is some consensus that SD-type programs appear to help adult women (especially labour force re-entrants), but are not especially beneficial for adult men or youth (Heckman, LaLonde, and Smith, 1999; Martin and Grubb, 2001). There is also some indication that training programs that serve to “signal” worker quality (by, for example, being targeted on those with higher skills or promising a valuable credential) may also be relatively successful. Whether evaluations of such training programs have adequately controlled for the effects of “creaming” remains a contentious point, however.

Table 3 provides a summary of the SD results for the six EBSM evaluations. In addition to recording these results in absolute terms (annual hours, earnings, and weeks of EI collected) the table also seeks to state these results relative to comparison group totals²⁶ so that they may be directly compared to other studies. Table 3 explicitly

²³ As discussed previously, such inaccuracies may have occurred either from alignment problems between the participant and comparison groups or from response errors in the survey data on the labour market history variables.

²⁴ Both statistically significant and insignificant results will be reported because, regardless of significance, the results represent the evaluations authors' choice for the best point estimate of the impact being estimated.

²⁵ This figure includes SD clients in apprenticeship programs also though this group was not generally included in most of the evaluations.

²⁶ Unfortunately, many of the evaluations did not provide comparison group means in their published reports. In order to construct the proportional outcomes therefore the following values were assumed: For active claimants: Annual hours – 1250; Annual Earnings -- \$20,000; Weeks of EI – 15. For former claimants: Annual hours – 1000; Annual Earnings -- \$16,000; Weeks of EI – 15. These figures

identifies results that are statistically different from zero²⁷, but reports the insignificant estimates also because they are still the best estimate of the program's impact.

In general the results in Table 3 for active claimants are encouraging. The typical gain in annual hours worked is about 100 hours and earnings gains are in the \$2,000-3,000 range. These figures are consistent with the presumption that most of the gains from SD came from increased employment, with a more modest increase in hourly earnings. In proportional terms the gains are quite large by international standards – about 10-15 percentage points in both hours and earnings. The picture for EI collections is more clouded. In part the ambiguity of the findings for EI may stem from the difficulty is differentiating between the reduction in EI collections that employment provides and the related renewed EI eligibility provided by the same employment – a topic we will take up later because the differential effects are more apparent in other interventions.

Although admittedly speculative, it might be useful to offer a few conjectures about why the results in Table 3 seem so positive and how some of the “outliers” in the table can be explained. As to the predominance of positive results, these may be in part explained by the focus of many SD interventions on obtaining credentials. A majority of SD participants report that their program provided some sort of credential for completion and there is empirical evidence that such credentials may serve as a signal about productivity to prospective employers (Martin and Grubb, 2001). A second reason for the preponderance of positive results may be the relatively weak labor markets that the comparison groups faced in some of the evaluations (such as those for British Columbia New Brunswick, and Newfoundland).

It is also possibly to offer some speculations about the outliers in Table 3. For example, the large earnings gains in Alberta may be in part attributable to the timing alignment difficulties in that evaluation discussed earlier. If comparison group members did not experience the “Ashenfelter dip” in their pre-program earnings, difference-in-difference estimates will make it appear that program participants had outsize earnings gains. It is difficult to assess the size of this effect without additional data analysis, however.

The negative earnings impact in Ontario also warrants an explanation. One possibility is that the worsening labour market in that province shown in Table 1 made the “lock-in” effect more salient. That is, Ontario participants ended their interventions just when jobs became less available. Members of the comparison group that had held onto their jobs during the downturn therefore had better earnings outcomes, at least over the short-term. Some evidence on that score is provided by the time pattern of the Ontario earnings effects which show a large improvement after the first post-program year (*Ontario Summative Evaluation*, Table A-2). Possibly the negative effect of lock in on the earnings of Ontario participants had largely disappeared after two years and subsequent earnings patterns were more in line with those in other provinces.

approximate the values given in the British Columbia and Ontario evaluations, but they may not be representative of those for comparison groups in other provinces.

²⁷ That is, the estimate is significantly different from zero at the .05 level on a two-tail test.

Sample size restrictions generally prevented the evaluations from estimating effects of SD separately for subgroups of participants. Three of the evaluations (British Columbia, Ontario and Quebec) do report that women had somewhat more favorable overall impacts on employment and earnings than men, though these results are not reported separately for SD participants only. It does seem likely, however, that the EBSM results mirror the international finding that women were more likely to benefit from training than men.

Estimated impacts of SD on former claimants are also shown in Table 3. In general these estimates were more variable than were the ones for active claimants. Still, the general pattern, with the major exception of results for British Columbia, appears modestly positive²⁸. The hours and employment gains in Quebec are especially large for former claimants. The authors of that evaluation attribute the results to the “enhanced employability” that participation provided to former claimants many of whom may have been out of the labour force for some time. For active claimants, all of whom had relatively recent labour market experience, this effect would have been much reduced. These observations suggest that the comparison group in Quebec may not have been especially well matched in the case of former claimants. But some problems in matching this group occurred in all of the evaluations.

Authors of the British Columbia evaluation do not offer a clear explanation for the large negative results for former claimants. They do point out that the reductions occurred “for all subgroups of former claimants based on client age, education, or participation duration.” But they do not explore whether these outcomes may have been influenced by the methodology employed. Specifically, although several of the other evaluations used EAS-only clients as a comparison group for former claimants, British Columbia did not follow this approach to measuring labour force attachment. Instead, they used a screening question in their survey to reject individuals who were “not unemployed around the start date of the participants they were matched to.” It is possible that this screen virtually guaranteed that the comparison group would have relatively large earnings gains making it appear that the participants had done poorly. Our discussions of the estimated impacts for some of the other interventions for former claimants in British Columbia offer further support for this possibility.

E. Results for Targeted Wage Subsidies (TWS)

TWS is the third largest Employment Benefit. Approximately 20,000 Canadians participated in target wage subsidy programs in 2005/2006 at a cost of \$97 million. These figures suggest that the wage subsidy provided to the typical client is about \$5,000 which would be roughly a 50 percent subsidy on a \$20,000/year job for six months²⁹. The program is therefore quite similar to the temporary wage subsidy program for on-the-

²⁸ In assessing the results it is important to keep in mind that the Saskatchewan evaluation did not estimate impacts separately for active and former claimants. Hence the results shown for that province represent an aggregated estimate.

²⁹ Although TWS subsidies may extend for up to 52 weeks (78 weeks for disabled workers) it appears that the typical subsidy examined in the evaluations lasted for about 4-6 months.

job-training that was part of the JTPA program in the United States. Relative to European experiences with wage subsidy programs, however, the TWS subsidy may offer a somewhat higher fraction of wages but for a shorter period than is typical. In both the United States and Europe wage subsidies are more explicitly targeted on specific subgroups of the population (usually low-skilled, younger workers) than is the case in Canada.

The theory of targeted wage subsidies is relatively complex. Of course, a general wage subsidy should increase both wages and employment. Katz (1996) provides a “guesstimate” that each 10 percent of subsidy should increase wages by about 6 percent and employment by about 2 percent. However, when a subsidy is targeted at only one class of workers, the situation becomes more complicated because the displacement of unsubsidized workers by subsidized ones becomes relevant. By some estimates such displacement effects may be as large as 80-90 percent.

The temporary nature of most wage subsidy programs adds further complications. In a simple economic model, the effect of a temporary subsidy should be smaller than for a permanent one because firms would not make the kinds of labour-using investments they might if the subsidy were permanent. But more complex models suggest that such differences will be less important when learning on the job is important. In these cases, the subsidy may help to compensate for an initial period of low productivity for new workers and help to overcome firms’ reluctance to make such hires.

These conceptual issues about wage subsidies have not played an important role in the EBSM evaluations, however, because the evaluations have focused exclusively on outcomes after the subsidy period ends. That is, the subsidy period is viewed as being the TWS “treatment” and this program is evaluated in ways similar to any other active labour market program. Reasons why a period of temporary subsidy should lead to future beneficial labour market outcomes are ambiguous, however. Certainly the end of a subsidy would make it more likely that a worker might be laid off thereby creating a negative effect relative to the comparison group. But it is also possible that skills and attitudes developed during the subsidy period may make employees more attractive

Despite this ambiguity, many evaluations of short-term subsidy programs have found beneficial outcomes. For example, the random-assignment JTPA evaluation in the United States found that female subsidy recipients experienced earnings gains of about 15 percent relative to the control group with males experiencing gains of about 10 percent (Bloom, et al. 1993). In many cases these gains persisted into the second post-program year. Similar positive results were found in the National Supported Work evaluation and in some of the analysis of some state’s welfare reform initiatives (Gueron and Pauly, 1991). Although evidence from formal evaluations is less readily available outside of the United States, a survey of OECD experiences offers the assessment that subsidy programs have a greater impact per dollar spent than either training programs or direct government job creation (Martin and Grubb, 2001).

One complication in conceptualizing the wage subsidy component of the EBSM program is in understanding precisely how it is targeted. Although most of the evaluations report that TWS is more appropriate for younger and harder to employ workers, few details are provided about how such targeting is achieved (for a further discussion, see Section V). The fact that the EBSM recipients studied in the evaluations all must have been EI claimants further complicates the targeting issue. In some respects EI claimants have characteristics more similar to dislocated workers than to the disadvantaged workers typically served by temporary wage subsidies. Precisely how the theory of wage subsidies apply to such workers is an open question. Similarly, the evidence on the effectiveness of such subsidies in achieving labour market gains among dislocated workers is much less well-developed³⁰.

Despite these caveats, the actual results reported for TWS in the evaluations were modestly promising, especially for former claimants. For example, Table 4 shows significant post-subsidy earnings gains for active claimants in two provinces (Alberta and British Columbia) and significant employment and earnings gains for former claimants in all of the evaluations that studied this intervention. In percentage terms, these impacts are roughly in line with the U.S. findings described previously (post-program gains of 15-20 percent in employment and earnings). The fact that TWS has a larger and more consistent effect for former claimants might have been expected. As described previously, this is a group for whom the rationale of a temporary wage subsidy may be the most applicable. Many former claimants have had substantial periods out of the labour force and the temporary subsidy reduces employer costs of getting them “up to speed” in their jobs. For active claimants, on the other hand, all have recent employment experiences, so the potential gains from TWS are not as great. These considerations suggest the importance of knowing precisely how TWS is “targeted” and whether that targeting could be improved – a topic we take up in several later sections.

The smaller (and sometimes negative) estimates for active claimants in TWS in New Brunswick, Newfoundland, and Ontario are consistent with the weak labour markets in those provinces. In this case, post-subsidy layoffs prove especially problematic for participants because they are unable to find new jobs. Many of those who do not remain with their subsidized employer therefore experience a negative “lock-in” effect because they have not been able to engage in the sort of exhaustive job search that may have characterized the comparison group.

As was the case for SD, the EI effects for TWS are mixed. They generally show a modest increase in benefits received relative to the comparison groups. This may in part reflect eligibility effects. Employment under a TWS plan is insurable under EI so eligibility is more-or-less automatic for most participants. Even if there are employment gains from the intervention it is still possible that some of those who lose their subsidized jobs will collect EI. Overall, however, the results from the evaluations suggest that this effect is not very large.

³⁰ Perhaps the closest analogy in the United States is to the re-employment bonus demonstrations. Although initial results from these demonstrations (in Illinois) were quite positive, results from the follow-on demonstrations were much less so (see Meyer, 1995).

F. Results for Self-employment (SE)

Self-employment provides financial assistance and other advice to help eligible participants start their own businesses. Total spending on SE amounted to about \$145 million in 2005/2006 with approximately 12,000 individuals starting this intervention during the year. Hence, on a per-participant basis this is a relatively high cost intervention – averaging over \$12,000 per client.

Provisions for self-employment have come to play an increasing role in the active labour market programs of many countries. In the United States Unemployment Insurance program individuals in some states may continue to receive their full benefits even if they are engaged in full-time self-employment activities³¹. Similar programs are available in most other OECD countries, although many of these are relatively small. Overall OECD countries spent only about two percent of total active labour market spending on self employment, so, by this standard the Canadian program is quite large.

The theory behind self-employment assistance is seemingly straightforward. It is well-known that most new job creation stems from small businesses and it is believed that some EI claimants may be effective at starting such enterprises. Relatively modest financial assistance can be used to overcome whatever entry barriers exist. Because claimants express strong preferences for “being their own bosses”, it seems that such assistance is a wise, if perhaps a bit risky, investment.

Research on self employment has raised some warnings about this scenario, however. The principle finding is that outcomes from spells of self employment are extremely heterogeneity. In some cases the ventures can be wildly successful, creating employment not only for the individual involved but for many others as well. In other cases spells of self employment can have serious negative consequences for the individuals involved – their businesses may be unsuccessful and they may incur a wage penalty when they seek to re-enter paid employment. A recent study of labour force dynamics finds that a one year spell of self-employment by men reduces subsequent market wages by between 3 and 10 percent (Bruce and Schuetze, 2004). Negative results were also found for women, but these were often not statistically significant because of small sample sizes for women entering self-employment. Despite these negative findings, however, the authors point out that the negative impacts from spells of self-employment are considerably smaller than those from unemployment itself. So, in the EBSM context, it may still be the case that self-employment is the best of the alternatives available.

The results for SE (Table 5) illustrate some of these ambiguities quite clearly. With the exception of Newfoundland, both active and former claimants in SE have significant increases in hours employed in all of the evaluations. Increases in annualized hours worked of 20-30 percent seem to have been the norm, with much larger gains being reported in Quebec. These strong gains suggest that many SE participants remain self-

³¹ These activities are reviewed in Vroman (1997).

employed after the formal intervention ends³² and that they generally report working full time on such jobs.

Unfortunately the employment gains from participating in SE were often not accompanied by increases in earnings – in some cases the evaluations report significant declines in earnings for SE participants. Such findings were not universal, however. Both Nova Scotia (for active claimants) and Quebec (for former claimants) reported significant increases in earnings once all sources of self-employment income were taken into account. Whether the differences between these positive findings and the other more negative ones can be explained by differences in ways in which the earnings data were collected is an open question³³. Clarifying whether SE provides a good income source for participants or, instead, raises difficulties in returning to paid employment should be an important goal of future evaluations.

SE participants generally experienced significant decreases in EI receipt in the post-program period. Because weeks in self-employment are not insurable under EI it is likely that these outcomes largely reflect eligibility effects rather than a decline in EI collection among eligible workers. If this decline were accompanied by declining earnings, the incomes of workers pursuing self-employment may experience serious declines, especially in the short run. Although this sort of impact was not found in all of the evaluations, the possibility that those in self-employment may experience large short-run declines in incomes suggests caution in expanding self-employment interventions beyond carefully targeted sub-groups of claimants.

SE participants generally experienced significant decreases in EI receipt in the post-program period. Because weeks in self-employment are not insurable under EI it is likely that these outcomes largely reflect eligibility effects rather than a decline in EI collection among eligible workers. Coupled with the declining earnings, this impact suggests that the incomes of workers pursuing self-employment may experience serious declines, especially in the short run. This in turn suggests caution in expanding self-employment interventions beyond carefully targeted sub-groups of claimants.

G. Results for Job Creation Partnerships (JCP)

Job Creation Partnerships are provided through community-developed projects. These jobs are intended to offer participants work experience and to benefit the local community. The number of JCP participants is relatively small – in 2005/2006 they constituted only about four percent of all new EBSM interventions and about five percent of total EBSM spending. These figures are smaller than at many OECD countries – where spending on public sector jobs amounts to about 15 percent of total active labour market expenditures. Such spending levels have been declining over time, however, in part because of relatively negative findings from evaluations of these types of programs. For example, Martin and Grubb (2001) conclude that such measures have been “of little

³² The evaluations that do report continued self-employment generally find that between 50 and 70 percent of participants in SE continue to be self-employed at the time of the survey (18-24 months post-program).

³³ Self-employment earnings may also suffer from under-reporting for tax and other reasons.

success in helping unemployed people get permanent jobs in the open labour market.” (page 32). They go on to point out that such jobs may have the temporary benefit of helping workers establish or re-establish connections to the labour market, but because (in these authors’ opinions) the jobs are “low productivity” they should only be for short durations. Some studies have reported that participation in public sector employment can help to improve training outcomes for low skilled workers. That is, the effects of training are more likely to stick for this group if they can experience a period of relatively undemanding work prior to joining the formal labour market (Heckman, LaLonde, and Smith, 1999). Interactions between these two programs were not explicitly studied in any of the EBSM evaluations, however.

Summary results for the JCP intervention are reported in Table 6. The most obvious conclusion to be drawn from the table is that the estimates are extremely varied. For active claimants relatively strong earnings gains in British Columbia are contradicted by the negative impacts in Ontario³⁴. This picture is reversed for former claimants in these two provinces – losses in British Columbia are balanced by gains in Ontario. Impacts on earnings for active claimants in Newfoundland and Quebec are essentially zero whereas for former claimants the Newfoundland evaluation shows losses whereas the Quebec evaluation shows earnings gains. Results for hours worked and for EI collections are similarly erratic.

Taken at face value, it is hard to make any sense out of these estimates. There is simply no consistent story to tell about the effects of participating in a JCP-type intervention. Of course, it may be the case that greater detail on how participants were selected for JCP projects or more information about the nature of the projects themselves would help to clarify matters. In Section V we examine whether the qualitative analyses from the evaluations can aid in this clarification. But we generally conclude that the analyses provided in the evaluations neither identify any good reasons for these differential outcomes, nor provide good evidence on the value of JCP projects to local communities. Hence, whether JCP intervention is an effective component of the EBSM program remains an open question.

H. Results for Employment Assistance Services (EAS)

EAS interventions focus on aiding job entry by participants. These services can include group activities such as job search workshops or access to job postings and individualized counseling including the development of Action Plans and referral to other EBSM interventions. EAS interventions serve by far the largest number of Canadians of any EBSM activity – in 2005/2006 approximately 430,000 individuals accessed EAS interventions. These services are available to all who wish to participate – in recent years somewhat more than half of the participants have been active or former EI claimants. EAS interventions are, on average, the least costly examined in the EBSM evaluations. Average costs in 2005/2006 were about \$1150 per participant. Most of the evaluations

³⁴ Again, the large gains for JCP in Saskatchewan may be an illusion because of the ways in which interventions and client groups were combined in that evaluation.

report average costs of EAS that are somewhat less than this – perhaps in the \$700-\$800 range.

Most of the evaluations have tended to treat EAS interventions as relatively uninteresting. The implicit notion seems to be that such low cost interventions would be unlikely to have impacts that can be measured given the constraints on statistical precision in the evaluations. This assumption is somewhat at variance with the evaluation literature which shows that in some cases the impact of such “minor” interventions can be relatively positive. For example, Meyer (1995) provides an extensive analysis of a number of experimental evaluations of job search assistance in the United States and concludes that these typically found reduction in unemployment durations in the 0.5-1 week range. Often such reduced unemployment was accompanied by increased earnings. The author does point out that there is some uncertainty about whether such results are caused by the actual services provided or perhaps by the increased monitoring that participants are under to ensure that they are pursuing employment actively. Martin and Grubb (2001) report similar findings for a few other OECD countries. Evidence on job search and related activities for youth in Canada also reaches relatively positive conclusions, though earnings effects are generally more ambiguous than in the other experimental studies (HRDC, 1997).

One complication in evaluating EAS in the EBSM context is that often these services are combined in action plans with other interventions. Perhaps the most extensive detailing of this is provided in *Annex C* of the British Columbia evaluation. There the authors show that 54 percent of all Employment Benefits in the province were accompanied by an EAS intervention and that in many cases the number of individual services received under EAS auspices was quite large. Because of this complexity, evaluations of EAS have tended to focus on the group of EAS “only” claimants. According to the British Columbia data, these represent perhaps 65 percent of individuals who received any EAS-related services, but a much smaller fraction of total EAS services provided (because those with an Employment Benefit intervention tend to have more EAS interventions than do the EAS-only group). The extent to which the EAS-only group is representative of all EAS participants has not been explicitly addressed in the evaluations, but on *a priori* ground it seems plausible that this group might have more successful employment experiences than the other EAS participants.

Table 7 presents the results for EAS-only participants. Only the results for active claimants are reported in the table³⁵. Former claimants in the EAS-only group were handled in a variety of ways in the evaluations and ultimately the results for this group are non-comparable. For example, the Alberta evaluation did not look at the EAS-only group at all and the Newfoundland and Ontario evaluations used EAS-only clients as their comparison groups for former clients in Employment Benefits interventions. For those provinces that did include an EAS-only group in their evaluations of former claimants, estimated results were highly variable and hard to interpret.

³⁵ Because Nova Scotia and Prince Edward Island used the EAS-only group as their comparison group for both active and former claimants, no results for these provinces are reported.

The results for active claimants shown in Table 7 are ambiguous. In two of the five evaluations reporting labour market impacts, the estimated hours and earnings impacts have opposite signs (though these estimates are usually not significantly different from zero). Estimated effects on EI collections are similarly erratic, with British Columbia and Quebec showing large increases. Overall then it seems that no overall conclusions can be drawn about the impact of EAS-only in the EBSM context.

A natural question is why these results seem to differ so much from the small, though relatively stable findings reported in many job search evaluations. Three possibilities might be mentioned. First, it may be that it is especially difficult to evaluate EAS using non-experimental methods (many of the best job search studies used random assignment). Measuring the impact of this low cost intervention may require a very precise matching of participant and comparison cases in the pre-program period and it may not be possible to achieve that precision with the propensity score methods used here. A second possibility is that the actual services delivered under EAS are more heterogeneous than the package of job search services usually studied. The fact that many claimants received numerous specific EAS services supports this view. Finally, many previous job search studies have been done in the United States where the provision of such services also plays a modest monitoring and enforcement role. That role may be less significant in Canada, and that may account for a reduced impact. Whatever the explanation, it seems that knowledge about EAS might be enhanced by a more explicit targeting of research effort on this activity.

I. Summing-up the impact estimates and some remaining questions

The impact estimates for the four Employment Benefit interventions seem generally consistent with results obtained in other studies. Specifically:

- Participation in Skills Development seems to have increased annual earnings by \$2,000-\$3,000 for active claimants – that is by about 10-15 percent (relative to the comparison group). Impacts for former claimants were less significant, but still predominately positive. The negative findings for active claimants in Ontario and for former claimants in British Columbia remain anomalies, however.
- Impacts of Target Wage Subsidies were most consistent for former claimants. One average this group achieved gains in earnings and hours worked of 15-20 percent relative to the comparison group. Results for active claimants were less consistently positive, however.
- Both active and former claimants in Self-Employment programs had large gains in hours worked (in the 20-30 percent range), but these increase were often not matched by increases in earnings. Participants in self-employment also experienced sharp reductions in EI collections, probably because such employment is not insurable under EI.
- There are no consistent findings from the evaluations of Job Creation Partnerships for either active or former claimants.

- EAS-only participants exhibited relatively small and statistically insignificant impacts in most of the evaluations' outcome variables. The effectiveness of employment services that were delivered in conjunction with major interventions was not examined in the evaluations, however.

The fact that these findings are broadly in accord with international evidence on active labour market programs lends support the approach taken in the EBSM evaluations. That is, the methodology chosen seems to have yielded believable findings, though a few findings are difficult to reconcile with this overall picture. Of course, many questions remain to be explored in assessing the validity of these findings. Here are five such questions that seem especially salient:

1. **Were the impact estimates robust to differing estimation techniques?** In the majority of evaluations only one version of the impact estimates was presented. In a few cases some alternative estimates are briefly mentioned, but mainly only to report that the results were “similar”. It is therefore difficult to tell whether the impact estimates would stand up to other estimation methods (i.e. using regression analysis rather than matching) or to minor variations in the specific methods used (using different matching algorithms or different regression specifications). Because estimates of the types reported in the evaluations can sometimes be quite sensitive to how they actually were obtained, it would add significantly to the perceived validity of the results if alternative techniques had yielded roughly similar results.
2. **Why did the anomalous impact estimates occur?** The evaluations tended to be quite modest in offering speculations about the reasons for their reported results. Such speculation might have been especially helpful in assessing what appear on the surface to be anomalous results. Further research in developing explanations for such results might be warranted.
3. **Will the impacts estimated persist?** Although the evaluations did illustrate some promising results, none of the interventions would pass a benefit-cost test if judged solely on the basis of the outcomes actually observed (see Section VI). Rather, gains in earnings or reductions in EI collections must be extrapolated into the future in order to cross such a hurdle. However, because the typical evaluation provided only 18-24 months of post-program experience³⁶, there is little basis for making such extrapolations. Similarly, in cases (such as SE) where short-term outcomes are ambiguous, only longer term data will clarify whether the intervention “works”.
4. **Did various subgroups respond differently to the EBSM interventions?** The analysis of subgroups in the evaluations was hindered by sample size considerations. A common approach was to estimate impact results separately by intervention, but to combine all interventions when looking at subgroups. Because the interventions worked very differently, such aggregation probably obscured

³⁶ Later evaluations had access to CRA data and these tended to provide a longer post-program period – in some cases up to four or even five years (Nova Scotia). In those evaluations with longer follow-up periods, earnings gains did not diminish greatly over time

whatever differences across subgroups may have existed³⁷. The most consistent finding in the international literature on active labour market programs with regard to subgroups is that female clients in job training programs tend to experience greater earnings gains than do male clients. There are some hints that this may also have been the case in the EBSM evaluations – the evaluations for both British Columbia and Ontario report that females had greater earnings gains than did males in both the active and former claimant categories (though these differences were not usually statistically significant). There are a number of other subgroups that might be of special policy interest, however. For example, low-skilled workers typically encounter severe labour market problems and it is common to target active programs at this group. Separate analyses for low-skill workers might therefore suggest how the implicit targeting of EBSM interventions might be improved. Similarly, it might be useful to look explicitly at seasonal workers to determine whether they might be affected by the EBSM interventions differently from those workers who are permanently separated from their jobs³⁸.

5. **Were variations within interventions important?** In all of the evaluations it was assumed that all of the participants in an intervention received essentially the same treatment. There are a number of reasons that this assumption might not hold: (1) Individual action plans often included more than one employment benefit; (2) All of the major EB interventions were accompanied by significant participation in various EAS activities; and (3) The resources devoted to participants within an intervention varied significantly. For example, costs of training programs varied across participants as did lengths of subsidized employment under TWS. Whether such differences mattered is not known, but could be important in the future structuring of action plans or in the targeting of interventions to specific groups of workers.

V. Qualitative Analysis of Interventions

All of the EBSM evaluations included extensive qualitative analyses. Data for these were drawn from the evaluation surveys, from key informant interviews, and from focus group discussions. This information was used to look at three broad questions:

1. How did EBSM activities actually operate in practice and what were participant experiences in them?
2. How did the EBSM program affect the operations of local labour markets – most importantly, what were the effects on employers?
3. How did the EBSM program affect local communities?

³⁷ The situation is complicated by the process by which claimants are assigned to interventions. If counselors are good at assigning members of claimant subgroups to interventions that will be most beneficial, it is possible that aggregated results may show important subgroup differences that might be obscured if such assignments were largely random.

³⁸ The issue of temporary versus permanent job separations was not examined in detail in the evaluations. Recent data from New Brunswick suggests that, in that province, a large fraction of EBSM participants are on temporary layoff and ultimately return to their prior employers. Probably most of these workers have jobs that are seasonal in nature.

In this section we look only at the first of these issues. Effects of the EBSM program on labour markets, employers, and communities are discussed separately in Appendix 3.

Ideally qualitative analysis can provide a valuable component to any evaluation. Perhaps the most important contribution is to provide a supplement to the quantitative impact analysis by clarifying precisely how the program under study was delivered and by suggesting useful hypotheses about the key determinants of program success or failure. This is the approach to be taken here. That is, we ask how the findings of the qualitative analyses can help to illuminate the impact results reported in the prior section – especially whether these findings can aid in understanding what worked and what didn't in delivering EBSM program components³⁹.

Unfortunately, the evaluations were sometimes not so explicit about the goals of their qualitative analyses. Rather than focusing on developing insights about the relative successes or failures of interventions, most of the analysis provided is relatively unfocused. Large portions are devoted to attitudinal questions about participant satisfaction⁴⁰ or to reporting relatively minor and individual-specific complaints. Rather than summarizing such results, this section is organized around six key issues about EBSM operations that could have been addressed in the qualitative analyses:

- How did Action Plans actually operate? How were participants assigned to interventions?
- What did participants in skills development actually do? How did the intensity of training programs differ across participants? What specific training goals were accomplished and how did these outcomes payoff in the labour market?
- Did participants in TWS jobs obtain significant skills that yield positive labour market outcomes in the future?
- What did participants in self-employment actually do? Were these activities likely to improve or worsen their future earnings prospects?
- What did participants in JCP do on their jobs? What kinds of skills were required and were those skills obtained?
- Which specific EAS activities did participants pursue? Which combinations of activities were most common, how intensive were they, and did participants gain from them?

³⁹ We do not examine EBSM eligibility criteria explicitly here. That is, we implicitly assume that targeting of active and former claimants is an appropriate goal for the program. Some significant shortcomings of this approach in cases where many potential beneficiaries are not EI-eligible are discussed in the Nunavut evaluation.

⁴⁰ We will not review the findings on satisfaction here. In general they report that large fractions of participants were “satisfied” with their interventions. Satisfaction is generally highest (75-80 percent) for SD and lowest for JCP and EAS (55-65 percent). In the absence of any comparative benchmark, it is hard to know what to make of these figures.

Answers to these questions are clearly necessary in order to interpret many of the impact findings. They are also important to drawing lessons about how the EBSM program might be improved in the future.

A. Action Plans

In theory the action plan concept is straightforward. EBSM participants, in consultation with their case managers, decide on a plan for increasing employability that includes access to one or more major intervention and the receipt of associated employment services. In practice, however, the action plan process is far from uniform. Not all provinces require the development of action plans and, even in cases where such plans are nominally required, the details can often be quite sketchy. One reason for these complications is that provincial data collection systems often do not include information on action plans *per se*, but instead focus primarily on collecting data on the timing of participation in specific intervention activities. In the evaluations, then, analysts have been required to construct “action plan equivalents” by using these intervention-specific dates together with decision rules that serve to define the action plan period and to identify the “primary” intervention for each participant. Given the complications involved in using these data it is not surprising that many participants could not provide any details on their action plans and significant numbers do not remember having one in the first place⁴¹.

Perhaps the primary loss from having such an imperfect understanding of the action plan process is that analysts have very little information about how EBSM participants were directed to specific interventions⁴². All of the interventions have descriptions that imply which subset of participants would benefit most from them and some of the evaluations provide data which show that demographic profiles do indeed differ across the interventions. But none of the evaluations provide much detail on how the allocation process worked in practice. This omission resulted not only in a loss of valuable contextual information about the ways EBSM programs operate, but it may also have missed an opportunity to collect data on participant assignment that would have been useful in the quantitative impact analysis. For example, Sochet and Burghardt (2007) use data from the U.S. Job Corps evaluation to show how information obtained from program intake staff about applicants’ program assignments and experiences can improve the quality of propensity score matching models, especially when these models are used to estimate program impacts on subgroups. It seems likely that similar improvements could have been obtained in the EBSM evaluations primarily by understanding how individuals were assigned to the smaller interventions such as TWS or SE.

B. Skills Development

⁴¹ In the British Columbia evaluation, for example, only 54 percent of participants report the development of an action plan as part of their EBSM activities.

⁴² In a recent paper Lechner and Smith (2007) question the value added by caseworkers in directing participants to the most beneficial treatments, however.

SD in the largest and one of the most costly of the interventions studied in the evaluations. It is also the intervention that yielded some of the most positive estimated impact results. Hence, a closer scrutiny of this intervention to see why it worked and how its operation might be improved is clearly warranted. Four questions seem especially important in this regard:

1. How did spending on SD vary across participants? Were training offerings largely homogeneous, or did some participants receive much more extensive training than others?
2. If training options varied significantly in content and expense, how were the costly slots allocated among participants? What impact did NFA (negotiated financial assistance) computations have on this process?
3. Did the intensity of training interventions payoff in terms of employment and earnings?
4. Even controlling for intensity, were some types of training more effective than others?

In general it is very difficult to answer any of these questions from the existing evaluation reports. For example, although all of the reports provide some data on the average costs of SD, none gives a clear picture of variations around this average. In some cases the reports do discuss the duration of time spent on SD and may even mention that these durations were quite variable. For example, the British Columbia evaluation reports that SD interventions averaged 159 days in duration, but their data show that the minimum duration was one day and the maximum was 1208 days. It is, of course, possible that these extreme values are incorrect. But it does seem that there was considerable variability in the duration of SD interventions across participants⁴³.

Because the evaluation reports generally did not document the variation in SD expenditures among participants, they also did not address how expensive training options were allocated among individuals. There is some indication that part of this allocation may have been achieved through price rationing in some provinces because of the ways in which NFAs were calculated. Several of the evaluations report participant dissatisfaction with this process, primarily because the rules for making the calculations did not seem clear. On average participant contributions were about \$1200-1500 and these constituted about 20 percent of training costs. How costs for more expensive training options were allocated is not discussed in sufficient detail to assess whether net pricing played an important role in allocating slots⁴⁴.

Some of the most common fields pursued by SD participants in the evaluations were health service training, computer skills development, and transport services.

⁴³ Further evidence on this is provided in the Newfoundland evaluation where it is shown that 2-year educational SD interventions were relatively common (though there was some confusion among participants about this availability).

⁴⁴ Other barriers to accessing expensive training options included the need for specialized equipment or an inability to attend training sessions on a full-time basis.

Intuitively it seems that these may indeed have been fields in which employment demand was growing⁴⁵, so those making training choices may have been fairly well-informed about economic realities. Whether specific training choices affected program impacts is not known, however.

Many of the evaluations report that most SD participants received some form of certificate or diploma from their training activities. Participants also reported that the skills they obtained from their training activities were important for obtaining their post-program jobs. These two facts may help to explain why SD was found to be relatively successful in the evaluations. On the other hand, certificates from training programs can sometimes be of dubious significance and respondent assessments are notoriously imperfect, so one should probably be cautious in drawing such conclusions.

Whether the intensity of training activities affected outcomes does not appear to have been tested in any of the evaluations. The primary reason for this is that most of the evaluations only identified individuals who participated in SD without augmenting their files with additional information about the extent of that participation. In some cases, such augmentation may have been impossible given the nature of the data available. Future evaluations would benefit from more extensive attention to the intensity issue.

C. Targeted Wage Subsidies

The stated goal of the TWS component of the EBSM program is to provide employers with financial assistance for wages of participants whom they would not normally hire without a subsidy. Placement in such jobs is expected to allow the participant to acquire skills and work experience that will enhance their labour market prospects once the subsidy ends. Given these presumptions, three questions might be asked about how TWS operated in practice:

1. How was TWS targeted so as to ensure that those who received the subsidy would not otherwise have been hired?
2. Did workers' experiences on subsidized jobs suggest that they were indeed acquiring useful skills? and
3. Did the skills obtained on subsidized jobs enable participants to get better jobs once the subsidy ended?

The evaluations' discussions of TWS targeting consist primarily of describing the characteristics of participants in this intervention. The most consistent finding is that these participants were more likely to be former claimants than active ones. This may indeed be indicative of a lack of recent labour market experience that TWS might address. There is also some indication that TWS were more likely to be recent immigrants or visible minorities (Ontario and Alberta) and the evaluations generally found that TWS

⁴⁵ The Ontario evaluation notes explicitly that some of the fields in which participants took training were affected by the economic slowdown in that province, however. This may in part explain the negative impact estimates for SD in Ontario.

recipients were a bit younger than other EBSM participants⁴⁶. Again, all of these pieces of evidence suggest that TWS may indeed have been targeted to the types of workers who required a period of subsidy to cover some initial on-the-job training.

Most TWS participants who were surveyed stayed with their employer until their subsidy ended. A majority also reported that they believed that this period of subsidized employment did provide them with training and experiences that should help them in other jobs. Specific skills acquired included enhanced abilities to work with computers, learning how to work in teams, and improved generalized problem-solving. Typically about three-fourths of TWS participants reported being quite satisfied with their experiences under the program.

The quantitative results suggest that TWS had a positive impact on post-subsidy employment and earnings, especially for former claimants. These results are consistent with the notion that TWS is being correctly targeted. But the evaluations shed little light on the precise elements that make TWS successful. Again, as was the case for SD, there is some evidence that the TWS intervention varied significantly in intensity. For example, the British Columbia evaluation found that although the average subsidized period was about 20 weeks, some much longer periods were recorded in the data. Whether these longer periods are simply mistakes in the data or actual program experiences that may pay off to participants is not known⁴⁷.

D. Self-employment

The self-employment intervention (SE) is perhaps the least well-understood of those provided under the EBSM program. In principle the intervention is targeted at those who appear to have a good business plan and the ability to implement that plan. Precisely how this assessment is made is unclear. Some indication of the implicit criteria being used is provided by the fact that SE participants tend to be older and better educated than other EBSM participants⁴⁸. SE participants were also more likely to be active than former claimants, suggesting that some weight may have been given to recent labour market experience in assessing whether self-employment would be a viable employment strategy.

By most accounts, SE is fairly successful in identifying individuals with viable business plans. All of the evaluation reports state that the vast majority of SE participants started their businesses as planned and more than half of these are still in operation at the date of the survey – typically more than two years after the start of self-employment. Many SE participants believe that the services provided to them were helpful in starting

⁴⁶ This was not the case in Newfoundland, however, where TWS participants were generally older.

⁴⁷ Whether the rate of subsidy provided under TWS varied across interventions is also unclear. All of the evaluation reports were rather vague about the actual fraction of wages subsidized under TWS. For example, the wording in the Ontario evaluation is typical: “ Under normal circumstances, the wage subsidy does not exceed 60% of the total wages paid to the individual for the period of the agreement.”

⁴⁸ In most of the evaluations females constituted a larger fraction of SE participants than they did of EBSM participants generally.

their businesses, though some believed that these services did not pay enough attention to market realities.

These reports of success stand in some contrast to the mixed results for SE reported in the impact estimates. Based on the quantitative analysis, it seems clear that SE participants were spending many hours working at their businesses, but the payoffs in earnings seem ambiguous. From the reported data it is hard to tell whether this absence of earnings is attributable to poor financial performance of participants' businesses or to earnings losses suffered primarily by individuals whose businesses have failed and are at a disadvantage when they must return to the formal labour market. The qualitative analyses do little to shed light on such issues. They provide only a very cursory study of SE participants' experiences with their businesses⁴⁹ and apparently did not explicitly study participants whose businesses failed. Hence, actual self-employment experiences of participants remain largely a "black box".

E. Operations of Jobs Creation Partnerships

The impact estimates for the JCP intervention were generally disappointing. These estimates suggest that participants acquired few valuable skills from participating in these programs. That possibility is modestly supported by the qualitative analysis in which many participants express some dissatisfaction with their JCP experiences. Two complaints were expressed in several of the evaluations. First, some participants expressed frustration that much of what happened during JCP was not geared toward the development of skills that would lead to long-term and sustainable employment. Second, many participants were dissatisfied with the fact that JCP jobs did not provide renewed EI eligibility as normal employment would have. Still, overall rates of satisfaction were relatively high among JCP participants and many expressed the view that they had indeed learned valuable skills such as time management or general confidence on the job.

In several of the evaluations participants reported that working on JCP jobs was a superior alternative to simply collecting EI. This insight provides an interesting alternative approach to studying JCP. Under this approach, attention might focus primarily on the program period and ask how the value of services produced by JCP projects compared to the opportunity costs involved in participating in those services. Unfortunately, both of these issues are difficult to address quantitatively. Measuring the value the "output" from JCP projects would require a significant research effort. Potential opportunity costs involved in participating in JCP were estimated in a few of the evaluations (these results are discussed in the next section), but the diversity of estimates for such costs was very large, ranging from about \$700 to over \$9,000. Similarly, none of the studies estimated whether participation in JCP had any impact on participants' job search activities. Given these uncertainties, it is unclear whether JCP participation really is a better alternative to simple EI collection for a segment of the low skill worker population.

⁴⁹ Several of the evaluations do report that SE participants expressed the need for greater "follow-up" in providing them with on-going business assistance – apparently indicating that some participants were experiencing problems in maintaining business viability.

F. Employment Assistance Services

The experiences of EBSM participants with respect to receipt of EAS interventions were extremely varied. Many participants appear to have received no such services⁵⁰ whereas some individuals are recorded as receiving over twenty interventions, some lasting several months. Given this heterogeneity, it is hard to know precisely what services EAS participants received. It is also difficult to determine how the services received by EAS-only participants compared to those who received similar services in conjunction with their participation in other interventions.

One reason that the data on EAS receipt are sketchy is that most employment services are provided by third-party vendors. This posed two problems for analysts. First, the types of services delivered were quite heterogeneous, making generalizations difficult. The most detailed documentation of services actually received is in the formative evaluations. These show that the most common services included under the EAS heading were client assessments, job search workshops, and employment-oriented group sessions that focused on labor market information and résumé preparation. But many participants also took brief courses on topics such as developing Internet skills or instructions on computers or other business machines. In some cases these courses extended for two or more weeks. For example, the Alberta formative evaluation notes that in some cases longer term (up to eight week) courses were included in EAS⁵¹. These focused on such career-related topics as chain saw safety, emergency first aid, and handling hazardous materials. Other provinces EAS programs may have included similar types of activities, though they may have in some cases been included under SD instead.

A second, related difficulty arising from the third party provision of EAS is the necessity of uploading data from service providers in a common format for HRSDC use. In most cases these data uploads indicated only participants identifying information and some data on intervention start and end dates. Details on services actually received were seldom provided. Hence, most quantitative information on the content of EAS received by specific participants in the formative evaluations comes from surveys, but these question batteries were generally not repeated in the summative evaluations.

The absence of detail on what EAS actually included for evaluation participants made it difficult to determine why the estimated impact results for this intervention were often so negative and contrary to other studies. Two possibilities are: (1) That some individuals were categorized as EAS recipients who actually received no services; and (2) That the EAS-only group may have been subject to selectivity problems that were not well-controlled through the matching process. With regard to the first possibility, the Saskatchewan evaluation found that large numbers of individuals who might have been termed “EAS participants” had only “requested counseling”. Apparently these

⁵⁰ This finding appears to be in part a matter of how receipt of EAS was defined – a definition that varied across and even within provinces. Whether participants actually could access major interventions without making any staff contacts is unlikely.

⁵¹ Alberta refers to its EAS-type services as “Career and Employment Assistance Services (CEAS)”. Saskatchewan uses a similar terminology.

individuals received no other services. If other provinces followed similar data collection procedures for EAS participants, it is possible that significant numbers who received no explicit services were included in the evaluation samples, thereby diluting the effects of this intervention⁵².

A second possibility is that selectivity issues were especially severe for the EAS-only group. For example, suppose those who sought EAS were similar to job-ready members of the comparison group along all measurable dimensions. But suppose also that these individuals had specific information about facing potential labour market difficulties that caused them to apply for services. In such a situation, it would appear that EAS caused these anticipated labour market problems. One piece of evidence that is consistent with this possibility occurred in the New Brunswick evaluation. There the authors found that far fewer of the EAS-only participants returned to seasonal jobs than did members of the “matched” comparison group. If participants sought EAS services because they knew that they would not be able to return to seasonal jobs (and this fact was not controlled for in the matching process), it would appear as if EAS “caused” this inability to become re-employed at a prior job. Because such seasonal effects were not extensively studied in the evaluations, such an effect may have occurred in other provinces.

VI. Analysis of program benefits and costs

All of the evaluations included some form of cost analysis, usually in the context of attempting to construct a simplified benefit/cost analysis of the EBSM interventions. In this section we look at the results of these studies. Before beginning, however, it may be useful to provide a brief discussion of some of the methodological issues.

A. Benefit-cost methodology

The methodology of benefit-cost analysis as applied to social programs has been the subject of extensive scrutiny over the past 40 years⁵³. The primary goal of this research has been to develop an accounting scheme under which “benefits” and “costs” are clearly defined from the varying perspectives of society, program participants, or the government (also referred to as “the rest of society”). Table 8 provides a simplified version of this accounting framework. Social benefits from active labour market programs include both the increase in pre-tax earnings received by participants (because these earnings reflect increased production of GDP) and the value of any non-monetary gains that may have been caused by the program (better children’s health, for example). Increased taxes represent a transfer from participant to the government, but net-out in a social accounting perspective. For similar reasons, any decrease in government transfer payments such as EI or SA do not enter into a social accounting (though they clearly do affect the government’s budget constraint).

⁵² Interestingly, Saskatchewan (where requests for counseling cases were excluded from the EAS group) had the largest estimated impacts of EAS on earnings and EI collections.

⁵³ For an extensive discussion of the issues discussed here, see Gramlich, 1990 or Boardman, et al., 2001.

The social costs of labour market programs have three components: (1) The resource costs of the government's provision of program services⁵⁴; (2) Any out-of-pocket costs that participants must incur to participate in the program; and (3) Opportunity costs associated with program participation. The third of these may warrant further discussion. When individuals participate in active labour market programs, they are losing the value of this time if it had been spent in some other activity instead. Specifically, participants lose the wages they might have earned had they not been involved with the program. This is clearly a cost both from the perspective of the individual (it will presumably affect whether they choose to participate in the program) and from the perspective of society as a whole (because less GDP will be produced while this person is in the program). Traditionally such opportunity costs can be measured by comparing the earnings of participants to the earnings of members of the control/comparison group during the program period⁵⁵.

The fact that program benefits and costs occur over time adds a further wrinkle to these calculations. Typically costs occur at the time of program participation, but benefits occur later, sometimes much later. This mismatch in timing often requires that two adjustments be made to benefits to make them commensurable with costs. First, benefits must be "discounted" back to the program period to allow for the "social rate of time preference". Usually this means that such benefits must be discounted by some sort of long-term interest rate on government bonds⁵⁶.

An assumption must be made about the duration of outcomes. Typically evaluations of active labour market programs cover only about two years of post-program outcomes. Hence, some assumption must be made about the extent to which measured gains persist into the future. Most researchers have found that observed gains tend to decay over time, sometimes at fairly rapid rates (see Heckman, et al. 1999). Such rates of decay can be incorporated into the analysis either explicitly by assuming benefits decline at an assumed rate (20-40 percent decay rates are typical) or by assuming that benefits last only for a specified number of years (3-5 years, say).

A final conceptual issue about making benefit/cost calculations is that often the estimates used for are subject to considerable uncertainty, primarily because they are based on relatively small random samples. Customarily authors take little account of these uncertainties. They use the point estimates from statistical estimates of benefits (and some costs) as the best available numbers and proceed to make the calculations described above⁵⁷. A more formal approach to the problem of uncertainty would devise

⁵⁴ Because paying for these costs must be done through compulsory taxation, it is also customary to include in program costs the "excess burden" of such taxes (typically 20-30 percent of resource costs).

⁵⁵ As discussed below, the computation is more complex for wage subsidy programs.

⁵⁶ There is some controversy about the proper interest rate to use to discount benefits of social programs with some authors claiming that the interest rate on government bonds is too high and others claiming it is too low. For a discussion in the Canadian context, see Burgess, 1981. The argument for discounting is unrelated to the issue of inflation. If benefits are measured in real terms, a real interest rate should be used in discounting. If benefits are measured in nominal terms, a nominal interest rate should be used.

⁵⁷ This is the procedure followed even in cases where estimated benefits turn out to be negative, because, assuming the estimates are consistent, this negative value is the best estimate of the program's impact.

decision rules that pay attention to the costs of both Type I errors (failing to reject an inefficient program) and Type II errors (rejecting a beneficial program). Such a decision-theoretic approach to benefit/cost analysis is uncommon, however.

B. Benefit/cost methods used in the evaluations

The benefit/cost sections⁵⁸ of the evaluations adopted a variety of approaches, paying varying attention to the issues discussed in the prior section. Some of the key issues that arose in the evaluations were:

- **Adopting a consistent perspective:** Usually benefit/cost analyses adopt a “social” perspective – that is they seek to know whether a particular program is a good investment from society’s point of view. At times analysts may also adopt a client-based perspective or a government budgetary perspective. But, as Table 8 shows, there is a clear logic to each of these perspectives – combining what are benefits from one perspective with what are benefits only under a different perspective may make it difficult to interpret the results. For example, some of the evaluations included EI or SA savings together with earnings gains as “benefits”. From a social perspective, reductions in EI or SA are not benefits. From a government perspective increases in earnings are benefits only to the extent that they yield added tax revenues.
- **Measuring only some costs:** All of the evaluations used administrative data to calculate per-participant intervention costs. A summary of these cost figures is provided in the next section. The evaluations were less consistent in their treatment of three other cost items, however:
 - Administrative costs of interventions;
 - Out-of-pocket costs for participants; and
 - Participant opportunity costs.

As Table 9 shows, practices on the degree to which such costs were estimated varied widely.

- **Making assumptions about discount and decay rates are not explicit:** Some of the evaluations include as benefits only those impacts actually observed during the post-program period. Others extrapolate these benefits using a variety of schemes. Discounting of future benefits is similarly subject to varied methods.
- **Statistical issues are not addressed:** Only the Ontario evaluation included statistical bounds on estimated benefits and these bounds were very wide (often including zero). None of the evaluations addressed issues involved in calculating Type I and Type II errors for their benefit/cost estimates.

Overall then the benefit/cost sections of the evaluations are difficult to summarize and provide little useful policy guidance. Section D seeks to develop a summary benefit/cost

⁵⁸ In some cases the evaluations provide a “cost-effectiveness” analysis which focuses on the ratio of program costs to estimated benefits. All of the conceptual issues discussed in connection with benefit/cost analysis apply to cost-effectiveness analysis also, though the latter tends to focus primarily on relative rankings among interventions whereas the former mainly asks whether any one intervention is a good investment.

analysis drawing from all of the data in the evaluations. First, however, we summarize some of the data on costs.

C. Cost estimates

Table 9 provides a summary of the cost figures used in the evaluations⁵⁹. The upper panel of the table shows the per-participant cost figures that each evaluation calculated using aggregate program costs attributed to an intervention during a particular fiscal year together divided by the number of participants starting interventions that year. Only in the Saskatchewan evaluation was an explicit attempt made to impute administrative costs to interventions – these costs averaged between \$320 and \$568 per intervention and are included in the figures for that province.

Perhaps the most interesting finding from the upper panel of Table 9 is the extent to which average intervention costs varied across jurisdictions. For SD, per participant costs were highest in Alberta, New Brunswick, and Quebec⁶⁰. TWS was especially costly in Alberta, Newfoundland, and Prince Edward Island. The calculated costs for SE and JCP in Ontario were especially large relative to the costs of these interventions in most other provinces. Whether these are true cost differences in delivering these interventions across the provinces or some artifact of the way in which the averages were computed is difficult to tell given the available data.

The middle panel of Table 9 reports the out-of-pocket participant costs estimated in the evaluations. Only the British Columbia evaluation provided a complete set of these cost figures. The data reported there seem reasonable. SE has the largest participant costs followed by SD. Participant out-of-pocket costs for the other interventions are relatively modest.

Estimated opportunity costs (foregone earnings) are reported in the bottom panel of Table 9. These figures were derived using the same methodology that the evaluations used to measure earnings impacts of the interventions with this methodology being applied during the within-program period⁶¹. Hence, the estimates should be treated with the same caveats and cautions as the impact estimates since they may be subject to the same sorts of selection biases. Still, the figures in Table 9 seem reasonable (with some exceptions) for the types of opportunity costs that participants incur when participating in EBSM interventions. The highest such costs are reported for SD, ranging⁶² from \$2700 to nearly \$7,000. Clearly participation in SD is a relatively time-intensive activity. Alternatively, the lowest opportunity costs are for EAS since using these services

⁵⁹ To simplify the presentation, only cost estimates for active claimants are presented. In general program costs did not differ significantly between active and former claimants, though there were some relatively large differences in estimated opportunity costs.

⁶⁰ Costs in Nunavut were about 20 percent higher than the larger figures in Table 9.

⁶¹ Evaluations that used CRA data did estimate differences between participant and comparison group earnings during the “program year”, but it is not possible to equate these with actual opportunity costs because they may contain both pre- and post-program differences.

⁶² The abnormally low estimate from Saskatchewan probably derives from problems with the intervention duration data in that evaluation.

requires only a modest time commitment. The finding of relatively low opportunity costs for TWS might also have been expected because participation in that intervention, by its nature, implies that the participant receives some employment earnings⁶³. The opportunity cost estimates for SE and JCP seem more problematic, however. For both of these interventions opportunity costs should probably be relatively high. In the case of SE, the initiation of a small business usually involves a substantial period of reduced earnings – a period during which the comparison group would be expected to have substantial earnings from employment. The finding may relate to how the “program period” is defined for SE, but the evaluations in which these opportunity costs were estimated provide few details. With respect to JCP, it seems likely that the opportunity cost computations were made from a client-based perspective. As Table 8 shows, from that perspective, the EI benefits received while working on JCP projects would count as a benefit whereas earnings lost by not having private sector employment would be a cost. In order to measure the social opportunity costs associated with JCP participation, researchers would have had to assign some value to the output produced by JCP projects. Opportunity costs would then be the difference between the value of the output produced by participants and what they might have produced in market employment (measured, say, by the earnings of the matched comparison group).

D. Illustrative benefit/cost estimates

Because of the large differences in methodology applied in the benefit/cost sections of the evaluations it is difficult to provide a succinct overall summary. Instead, in this section we provide some illustrative estimates for each of the interventions based on “typical” benefit and cost patterns summarized previously. These illustrations provide a general picture of the relative success of the various EBSM interventions and pinpoint some of the key elements leading to such an assessment. All of our illustrations are based on quantitative results for active claimants. Cases where the results would be significantly different for former claimants are discussed only in passing.

1. Skills Development: SD is the prototype active labour market program, so the methodology for assessing the benefits and costs for this intervention is well-developed. Table 10 provides an illustration for active claimants⁶⁴. Consider first the social perspective. In this case the only benefit measured here is the increase in post-program earnings. Based on the summary in section 3, we assume that the gain in annual earnings is \$2,250 – about the midpoint of the estimated range reported in Table 3. Assuming that this gain is subject to an annual decay rate of 20 percent (a relatively optimistic assumption⁶⁵) and that future benefits are discounted at a rate of 5 percent, the total discounted earnings gain is \$9,000 ($=\$2250/0.25$). Social costs of SD consist of about

⁶³ Estimated opportunity costs for TWS in British Columbia represent those for “all other interventions”, though the authors do not explain why they made the calculation in this way. Complexities in making benefit/cost estimates for TWS are discussed in detail below.

⁶⁴ For former claimants in SD representative estimates would be quite similar.

⁶⁵ The assumption of a 20 percent decay rate is mathematically identical to the assumption that earnings gains last five years.

\$6,000 in operational and administrative costs⁶⁶, \$2,000 in out-of-pocket costs for participants, and \$4,000 in opportunity costs. Based on these assumptions, the costs of SD exceed the benefits by about \$3,000 per participant. Given the uncertainties associated with all of the assumptions made, a reasonable summary would be that SD comes fairly close to being a wash from a social point of view. Benefits might exceed costs if earnings gains proved to be significantly greater in the future than assumed here or if there were significant social benefits other than earnings that accrue to SD participants.

Two additional components are required to calculate benefits and costs from participant or government perspectives. First, we assume that earnings gains or losses experienced by SD participants are subject to a 25 percent rate of taxation. Hence, the \$9,000 gain in earnings provides a net return of \$6,750 to participants and \$2,250 to the government. Similarly, foregone earnings of \$4,000 would have incurred \$1,000 in taxes and this figure enters into the calculations from participant and government perspectives (though not from the social perspective). Second, although the EI results reported in Table 3 are quite varied, we assume annual EI savings⁶⁷ of about \$250 (that is, about one week). Applying the 20 percent assumed decay rate together with a 5 percent discount rate yields a present value of EI savings of \$1,000. Given these assumptions, Table 10 shows that participants in SD derived a small gain from their participation. Since participation is voluntary, this result might have been expected. From the government's perspective, SD has a net cost that is about 60 percent of what is actually spent on the program. That is, roughly 40 percent of the costs are returned in increased taxes or reduced EI payments. Of course, all of these calculations are for illustrative purposes only. Although they are consistent with the general findings of the evaluations, there remains uncertainties about the true values of most of the benefits and costs included in the examples.

2. Targeted wage subsidies: Table 11 provides an illustrative benefit/cost analysis for TWS. As before, we base the estimates on figures for active claimants. Although estimated earnings impacts for this group were varied, a consensus estimate might be that these participants had annual post-program earnings gains of about \$2,000. As discussed previously, gains for former claimants were about 50 percent larger than this, so for them the benefit/cost analysis would be more favorable. If we again use a 20 percent decay rate and a 5 percent discount rate, the discounted value of the post-program earnings gains from active claimants in TWS is approximately \$8,000. This is the social benefit of the program⁶⁸. Calculating costs for TWS is conceptually complex. From a social point of view, the wage subsidies provided under the program (which averaged about \$4,500 per participant or which \$1,125 was returned to the government in increased taxation) are not a "cost" of the program. These payments are transfers. They enter as positive benefits from the participant's perspective and a cost from the

⁶⁶ Because these calculations are for illustrative purposes only, welfare costs associated with the taxes necessary to finance EBSM interventions have not been included in the cost figures.

⁶⁷ In these illustrative calculations we make no allowance for SA impacts.

⁶⁸ Because the estimated effects of TWS on subsequent EI collections were quite varied, the effect is estimated as zero in Table 11.

government's perspective, but cancel out from the perspective of society as a whole. Only the administrative costs of TWS (estimated at \$500 per participant) are true social costs. Social costs of TWS also consist of any out-of-pocket costs for participants (the British Columbia evaluation estimated these at about \$650) plus any opportunity costs.

To evaluate the opportunity costs associated with TWS, we must ask the counterfactual question, "What would GDP have been in the absence of the program?" That is, the value of the net increase in production on subsidized jobs must be weighed against the reduction in value of production that would have occurred if those in subsidized jobs had been employed elsewhere. Making this calculation is no easy task. One common assumption in benefit/cost analysis is that the overall level of employment is more-or-less fixed by macroeconomic conditions. Under this view, wage subsidies encourage workers to take lower productivity jobs than they ordinarily would have, so there is some net reduction in GDP. In Table 11 we rather arbitrarily assume that this reduction amounts to \$1,000 per participant⁶⁹ on which taxes of \$250 would have been collected.

Assembling all of these numbers shows that TWS had quite significant social benefits for active claimants – nearly \$6,000 per participant. This result occurred primarily because the relatively large post-program earnings gains of participants were not offset to any major degree by the social costs of the program. Of course, the calculation might have been quite different if larger opportunity costs for TWS had been assumed.

The remaining calculations in Table 11 make clear that participants themselves may benefit significantly from TWS. Most of those gains occur in the post-program period in the scenario illustrated here. Government costs of the program are relatively modest in these calculations, largely because of the assumed tax revenues generated by post-program earnings gains. The costs to the government would be significantly larger if these post-program tax revenues failed to materialize.

3. Self-employment: Illustrating a benefit/cost analysis for SE also involves a variety of conceptual and empirical challenges. With regard to benefits, the impact results found little evidence to support the hypothesis that SE participants experience post-program earnings gains. A consensus estimate⁷⁰ might be that these participants suffer annual earnings losses of about \$2,000, though there is considerable variability around this figure. The impact results also found significant reductions in EI collections by SE participants, amounting to perhaps five weeks of benefits (say \$1250 on an annual basis). The present values of these impacts (\$8,000 and \$5,000 respectively) significantly affect the benefit/cost calculations in Table 12.

Most of the evaluations conclude that SE is a rather costly intervention (see Table 9). From a benefit/cost perspective, however, it is necessary to differentiate between those aspects of cost that are transfers⁷¹ and those that are real resource costs associated

⁶⁹ In the British Columbia evaluation the authors used a similar figure for the opportunity cost of TWS using the rationale that this was the average opportunity cost of all interventions other than TWS.

⁷⁰ It should be kept in mind that some self-employment earnings may not be reported.

⁷¹ The *Monitoring and Assessment Report* (HRSDC, 2007) states that the financial assistance provided by SE is, in part, intended to cover recipients' living expenses while they act to establish their businesses.

with starting of a self-employed business. In general the evaluations did not provide the kind of detailed data that would permit making this differentiation. For purposes of the illustration in Table 12, we simply assume that the average cost of the SE intervention is about \$10,000. \$2,000 of this represents spending on real resources associated with business start-up, \$500 represents pure administrative costs, and \$7,500 represents transfer payments (primarily in-program EI collections).

Out-of-pocket costs and opportunity costs for SE were also found to be relatively large in the evaluations. Here we use the out-of-pocket estimate from British Columbia (\$4,600) together with an average of the Ontario and British Columbia opportunity cost estimates (\$2,600). These estimates of opportunity costs seem relatively low considering the time commitment usually involved in setting up a business, but these are the only data available.

Pulling all of these figures together provides a bottom line this is not favorable to SE. From a social perspective, per participant costs of the program outweigh the benefits by over \$18,000. These costs are shared fairly evenly by participants and the government. The only significant gain to participants is the receipt of living expenses while establishing the business whereas the primary gains to the government's budget are EI savings. Of course, this negative assessment is highly contingent on the absence of significant earnings from self-employed activities. If these earnings from self-employment prove eventually to be rather substantial (as they were in the Nova Scotia evaluation) this negative benefit-cost assessment could be reversed.

4. Job Creation Partnerships: Developing an illustrative benefit/cost analysis for the JCP intervention also presents difficulties. The most important of these is the need to have some estimate for the value of the output being produced in JCP activities. Usually in benefit/cost analysis the value of output is taken to be accurately measured by the earnings of workers. But for JCP that is clearly not the case. The "wages" that JCP recipients receive are almost exclusively EI benefits and should be regarded as transfers. If JCP projects produce useful goods and services, these clearly should be valued – but it cannot be presumed that these values bear any necessary relationship to the value of EI benefits received.

For illustrative purposes we assume that the JCP participants produce output worth \$4,000 per participant. This value should be weighed against the resource cost involved in JCP projects to assess the net social costs of such projects. According to Table 9, JCP costs about \$6,500 per participant. But a large portion of that "cost" represents EI transfers, not true social costs. In the absence of any good information on this subject, we assume that \$2,000 of JCP costs are true resource costs. These, together with an assumed \$500 of administrative costs, yield a net social benefit of \$1,500 per JCP participant (\$4,000 output value less \$2,500 in operating and administrative costs). In Table 13, this figure is shown as a positive value in the "cost" calculation (that is, as a gain rather than a cost). The remaining \$4,000 in per participant operating costs for JCP is assumed to be an EI transfer.

Annual post-program earnings gains by JCP participants are assumed to be \$1,000 – a relatively optimistic figure in the middle of the largely divergent estimates in Table 6. EI savings are taken to be about one week (\$250). Hence, the discounted values of these two impacts are \$4,000 and \$1,000 respectively. Finally, opportunity costs associated

with JCP participation are assumed to be about \$3,000. These costs represent what JCP participants might have earned on private sector jobs. The low figure for this estimate is consistent with the presumption that many JCP participants would probably have been unemployed had it not been for the program. The fact that our estimate for the value of output produced on the JCP job exceeds this opportunity cost reflects the possibility that these participants would indeed be more productive in public employment.

Given all of these assumptions, JCP looks relatively attractive from a social point of view. The program is also attractive to participants and only modestly costly from the government's perspective. Of course all such calculations are highly conjectural – they are dependent on the key assumptions made in Table 13 about the value of JCP output, moderately beneficial impacts, how JCP costs should be divided between transfer and resource costs, and the low assumed value for participant opportunity costs. Still, the calculations may explain why JCP continues to be a relatively popular option in the EBSM inventory despite the negative findings of most evaluations of public employment programs.

5. Employment Assistance Services: As discussed in section 4, the impact results for EAS were quite varied and followed no consistent pattern. Hence, for illustrative purposes, we have included a value of zero for all impacts of these services. As a result, the services only have costs reported in Table 14. In general these costs are modest – amounting to \$1,550 from a social perspective. It is clear therefore that even modest beneficial outcomes would make EAS a good social investment. But such beneficial results generally were not found for active claimants⁷².

VII. Lessons Learned

The EBSM evaluations provide a huge amount of information about Canadian active labour market programs. In this section we first seek to bring together all of the material surveyed previously to draw some conclusions about policy. The second part of the section is devoted to exploring the lessons that these evaluations may have for the conduct of future labour market evaluations.

A. Lessons for Policy

The general conclusion of the EBSM evaluations is that these active labour market programs have had a mixed record in achieving their goals of “enhancing the skills Canadians need to prepare for, obtain, and maintain employment.” In this section we ask whether the evaluations provide any insights about how this record might be improved. Before starting the discussion of specific interventions, a more general point about the participants in these programs should be made. Throughout the evaluations, the sets of participants in interventions have been taken as given by some form of allocation process under which participant desires combine with staff advice to determine which interventions will be pursued. The evaluations did not cast much light on how this allocation process worked in practice nor about whether alternative targeting mechanisms

⁷² Too few results were reported for former claimants to permit any analysis.

might have worked better⁷³. Similarly, because of sample size constraints, the evaluations did not provide the sort of detailed sub-group analysis that might have provided guidance to program staff about how interventions might be better targeted to those who would gain the most from them. Hence, although the “Action Plan” concept seems to be widely accepted as being a good way to address the needs of EBSM participants, it may not be achieving its full potential because of the limited information on which assignment to interventions is based. Clearly some additional research on this process seems warranted.

With regard to the specific interventions in the EBSM package, the evaluations offered a number of policy-relevant insights:

1. The actual content of Skills Development may be crucial

The SD intervention seemed relatively successful in generating earnings gains, though the program probably fell a bit short of meeting a strict benefit/cost test. Generally it appeared that experiences under SD were consistent with findings from other studies both in the magnitude of earnings gains estimated and in the fact that these gains seemed larger in relatively weak labour markets. The fact that SD seemed to perform less well in weakening labour markets suggests that some attention should be paid to the content of training programs in such a context and whether they meet the changing needs of employers. The intensity of training programs is also a potentially important issue. Our illustrative benefit/cost calculations in the previous section make clear that program costs play a pivotal role in determining whether SD “pays off” from either a social or a governmental perspective. Phrased another way, it would be useful to know whether more expensive training options can be justified by the larger potential earnings gains generated. Finally, the role of out-of-pocket costs associated with SD participation may be an important issue. Those evaluations that tried to measure these costs found them to be relatively large – approximately \$2,000 per participant. And a number of evaluations reported that participants found the process of negotiating their financial contributions to be opaque. Whether these factors influence participation in SD or affect which kinds of training are pursued is not known. But because carefully structured co-payment schemes may have an important influence on training choices, it is important to know more about the effects of this potential policy tool.

2. The targeting of Targeted Wage Subsidies is important

The TWS intervention had significant positive effects on earnings in many of the evaluations and the analysis of the previous section suggests that these benefits may exceed the program’s social costs. Impacts for TWS seem especially strong for former claimants, perhaps because these individuals need a period of subsidized employment to re-adjust to the labour market. The finding the TWS may work better for former claimants suggests the potential importance of the way in which this intervention is in fact “targeted”. Although descriptions of TWS stress that they are intended for workers who would “not ordinarily be hired in the absence of a subsidy”, evidence from the

⁷³ For example, Lechner and Smith (2007) suggest that statistical decision rules may do a better job of assigning clients to services than do caseworkers.

evaluations on whether this is indeed the case is sketchy. In most provinces TWS has more former claimants than active ones and some demographic characteristics (such as that TWS participants are somewhat more likely to be female than participants in other interventions) also suggest that some targeting toward those with less recent labour market experience may be occurring. But differences among demographic groups are not large and in some cases seem inconsistent with active targeting (for example, the finding that educational differences for TWS participants do not differ from those in other interventions).

Related to the targeting issue is the question of the “displacement” caused by wage subsidies. If TWS actually creates new jobs for workers who would have otherwise been unemployed, the output from these jobs is a social benefit of the subsidy⁷⁴. Alternatively, if the subsidy just determines who gets a job that would exist anyway, there are no such gains. None of the evaluations sought to estimate this displacement effect in any detail and this absence shows up in the complex nature of the discussion of the opportunity costs of TWS participation. In our illustrative benefit/cost analysis we assumed that displacement was complete – that is, no new jobs were generated by the TWS subsidy. Some of the anecdotal evidence from the evaluations is consistent with this assumption – employers often report that they would not have hired TWS participants in the absence of a subsidy (they would have hired someone else instead). But this question is far from settled.

3. The social gains from Self Employment need to be clarified

The SE intervention did not show up well in many of the evaluations. Participation in this program appeared to be quite costly from all three benefit/cost perspectives and evidence of long term self-employed earnings to offset these costs was meager. On the other hand, all of the evaluations found that relatively high fractions (over half) of participants in SE were continuing to operate their businesses at the time of the survey, perhaps two years after obtaining the help that SE provided. The most important evaluation need, therefore, is to clarify the welfare effects of the self-employment promoted by this intervention. The social benefits of self-employment need not stem only from earnings. It may also be the case that self-employment provides many non-monetary benefits such as better working conditions or hours flexibility. But these benefits need to be documented – especially in the light of the many extra hours that the self-employed report working.

The evaluations provided some evidence that SE is carefully targeted toward those with viable plans for self-employment. Those pursuing this option were also found to be somewhat older and had higher levels of education than other EBSM participants. Still, there were some comments from the surveys and focus groups expressing dissatisfaction with the sort of market information that staff can provide to SE participants. Although staff members were helpful in many of the accounting and permitting issues associated with self-employment, they apparently were less so in

⁷⁴ The evaluations measured impacts of TWS on *post-program* employment and earnings where there is no ambiguity about the social value of earnings gains

assessing whether businesses were viable in a financial sense. Of course, making such assessments is no easy task – the uncertainties in evaluating a business plan are pervasive. Still, it may be useful to bring additional expertise about marketing and financial viability into the SE process.

4. Who Participates in Job Creation Partnerships and what they do matters

The impacts estimated for JCP participants were quite varied in the evaluations. In some provinces relatively large post-program earnings gains were reported, especially for former claimants. In other provinces, however, estimated returns from JCP were negative, sometimes significantly so. One possible explanation for such varied outcomes is in variations in the way the JCP is targeted. Most of the evaluations reported that JCP participants did indeed have demographic profiles that suggested that they may have required some assistance in gaining labour market experience. But it is possible that some provinces had stricter screening procedures for participation than did others. Clarifying these procedures would be an important step in understanding the determinants of JCP success.

The illustrative benefit/cost calculations in section 6 also make the point that any overall assessment of JCP requires that some estimate be made of the value of output from JCP jobs. Although some of the analyses of the community impacts of JCP did mention that the projects had “great value” to the communities in which they occurred, specific details on the nature of this value were often lacking. And the fact that many participants noted that their JCP jobs seemed to provide few skills relevant to private sector employment suggests caution in accepting such qualitative assessments at face value.

5. The EAS-only results remain a puzzle

Participation in EAS interventions was by far the most common activity in all of the evaluations. This occurred both together with participation in employment benefit interventions and in isolation. In general there were no attempts in the evaluations to study how EAS packages may have fostered the goals of the employment benefit interventions, but the EAS-only group (for active claimants) was examined in most of the evaluations. The negative findings for this group remain a puzzle. Most random assignment evaluations of such “minor” interventions conclude that they have small, but statistically significant effects on collection of unemployment insurance benefits and occasionally effects on earnings have also been found. It may be that EAS interventions in Canada do not play the same rules enforcement role that they play in the United States (where most of the random assignment studies have been done) so that this component of the observed effect is not operative. It may also be that the matching methodologies used in the evaluations are particularly difficult to implement in the EAS-only case. Whatever the underlying reason, because significant resources are expended on EAS activities, further empirical research would be helpful in identifying the most successful components of this complex set of programs.

B. Lessons for Evaluations

Experiences with the EBSM evaluations also provide a variety of lessons about the conduct of future quantitative labour market assessments.

1. The methodologies used in the evaluations seem to have been generally successful, though some problems in validation remain.

Many of the evaluations yielded sensible impact estimates and this adds some confirmation of the success of the measurement methods chosen. It was recognized at the outset that the matching methods adopted were generally less robust to potential problems of self selection than a purely random assignment would have been, though they were probably the best that could be done given the constraints under which the evaluations operated. Researchers generally seem to have used “state-of-the-art” methods for matching and usually reported that statistical tests for the quality of such matches were supportive.

Still, some problems remain in assessing the quality of the methods used. In part these problems are inherent in matching. One can always question whether the right variables were used or whether unmeasured factors were especially important in specific instances (as appears to be the case with EAS). Perhaps the two most important potential problems in controlling for measurable differences in matching were: (1) Aligning the timing of the participant and comparison groups so as to control for the pre-program dip in earnings usually experienced by program participants; and (2) Controlling for seasonal factors in employment and the problem of returning to prior jobs. Future evaluations based on matching should pay special attention to these two factors.

2. The quality of survey data may warrant additional examination

Most of the evaluations reported relatively low response rates on surveys and found that the quality of responses to some items (notably those on labour market history) was problematic. Such findings, in combination with the increasing availability of CRA data for evaluation purposes, resulted in a reduction in reliance on survey data in the more recent evaluations. This trend certainly poses some advantages in evaluations. Using administrative data is a low-cost way to increase sample sizes, and worries about non-response bias in such data sets is minimal. But the use of administrative data, especially CRA data on earnings, is not a panacea. Often administrative data do not include specific outcomes of interest (for example data on hours of work) and the time aggregation used in them (the CRA data are annual) may obscure important short term effects. In addition, administrative data sets provide little contextual information about participant’s specific economic situations and about their interactions with programs. Because surveys can in principle fill these voids, some attention should be devoted to developing ways to improve the quality of the data they yield.

In this regard, it would be useful to develop additional studies of the quality of the survey data, perhaps using CRA data as a benchmark. This should pinpoint areas in which survey data are especially problematic. Such information could be combined with

details about what the survey efforts in the evaluations cost (and what fraction of those costs were devoted to data quality issues) to see whether added spending in some areas might payoff in terms of increased quality. Comparisons to survey efforts elsewhere might also help to illuminate whether the Canadian evaluations seem to have an inordinate number of survey quality issues.

3. The evaluations would have benefited from a greater degree of uniformity in the presentation of results and from the availability of public use data files.

Employing multiple contractors in for the EBSM evaluations offered the opportunity to benefit from a wide range of expertise. A drawback of this approach, however, was that there was too little uniformity in the way that the impact results from the evaluations were presented. Outcome variables were not defined consistently, interventions were sometimes combined in unusual ways for estimation, and there was no consistent treatment of subgroup estimation. This situation made it difficult to compare results across evaluations or to draw summary conclusions about the efficacy of specific interventions. Contractors clearly should be given considerable flexibility in the design and analysis of evaluations, but some thought might be given to also requiring a set of core results to be reported in a uniform manner across an entire set of evaluations.

One way to achieve uniformity in results would be to require that evaluations provide public use files so that researchers can combine data across a set of evaluations in any way that seems desirable. The creation of such files would again require some uniformity in variable definitions, but such issues as the combining of interventions or conducting subgroup analyses could be addressed in any way that a researcher thought appropriate. A potential objection to the creation of public use files concerns the need to preserve data confidentiality. But devising procedures to achieve this end should not be especially difficult.

4. Qualitative analysis in the evaluations needs to be focused more explicitly on policy questions about intervention delivery

All of the evaluations devoted significant resources to the development of various types of qualitative analyses. Although these did provide some interesting insights, overall they fell a bit short of providing a comprehensive picture of how EBSM operated. Hence, they contributed only marginally to providing a context in which to view the impact results. Nor did they identify key policy parameters or processes where changes might significantly improve chances for program success. Future evaluations should seek to integrate the impact and qualitative analyses more closely.

The evaluations also used their qualitative analyses to assess the employer and community impacts of EBSM. Generally these analyses posed a number of practical and conceptual challenges, primarily because the data in the evaluations were not adequate for such purposes. If future evaluations are to make serious attempts at measuring employer and community impacts, more precise (and more expensive) methodologies will be necessary.

5. How benefit/cost analysis is included in the EBSM evaluations might be reconsidered

Developing comprehensive benefit-cost analyses can be a difficult and expensive proposition in any evaluation. Careful attention must be paid to such questions as: (1) developing consistent perspectives (participant, government, or social); (2) measuring all costs, including opportunity costs; and (3) addressing a number of technical issues related to how benefits and costs are aggregated. Of course, creating such a full-blown benefit/cost analyses can be quite expensive (consider the costs of measuring the value of output on JCP jobs, for example). But reporting partial benefit/cost analyses can be misleading in assessing the desirability of a project. Hence, it may not be a good strategy to include benefit/cost components in all evaluations. Instead, such analyses might be contracted separately or included as components only in those evaluations where their results might be most useful. Because benefit/cost analyses of EBSM programs is an important component of governmental accountability initiatives, more effective strategies need to be considered for developing them in a comprehensive way.

References

- Ashenfelter, Orley. (1978)** “Estimating the Effects of Training Programs on Earnings” *Review of Economics and Statistics* January, pp. 47-57.
- Bloom, Howard, et al. (1993)** “The National JTPA Study Overview: Impacts, Benefits, and Costs of Title II-A.” Report to the U.S. Department of Labor, Abt Associates.
- Boardman, Anthony E., Greenberg, David H, Vining, AidanR., and Weimer, Donald L. (2001)** *Cost-Benefit Analysis: Concepts and Practice, Second Edition*. Upper Saddle River (NJ): Prentice Hall.
- Bruce, Donald and Schuetze, Herbert J. (2004)** “The labor market consequences of experiences in self-employment.” *Labour Economics*, 11, pp. 575-598.
- Burgess, David F. (1981)** “The Social Discount Rate in Canada: Theory and Evidence.” *Canadian Public Policy*, VII, pp. 383-394.
- Gramlich, Edward M. (1990)** *Cost Benefit Analysis, 2nd Edition*. Englewood Cliffs (NJ). Prentice Hall.
- Gueron, Judith and Pauly, Edward (1991)** *From Welfare to Work*. New York. Russell Sage Foundation.
- Heckman, James; LaLonde, Robert and Smith, Jeffrey. (1999)** “The Economics and Econometrics of Active Labor Market Programs,” in Orley Ashenfelter and David Card, eds., *Handbook of labor economics*, Vol. 3A. Amsterdam: North-Holland, pp. 1865-2097.
- Heckman, James and Smith, Jeffrey. (1999)** “The Pre-programme Earnings Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies.” *Economic Journal*, 109, pp. 313-48.
- HRDC. (1997)** “Lessons Learned: Effectiveness of Employment-related Programmes for Youth.” Ottawa. Human Resources Development Canada.
- Katz, Lawrence F. (1996)** “Wage Subsidies for the Disadvantaged” *NBER Working Paper 5679*. Cambridge (MA) National Bureau of Economic Research.
- Lechner, Michael and Smith, Jeffrey. (2007)** “What is the value added by caseworkers?” *Labour Economics*, 14, pp.135-151.
- Lechner, Michael and Wunsch, Conny. (2006)** “Are Training Programs More Effective When Unemployment is High?” Bonn. *IZA Discussion Paper 2355*.

- Martin, John P. and Grubb, David. (2001)** “What works and for whom: A review of OECD countries’ experiences with active labour market policies.” *Swedish Economic Policy Review*, 8, pp. 9-56.
- Meyer, Bruce. (1995)** “Lessons from the U.S. Unemployment Insurance Experiments.” *Journal of Economic Literature*, Vol. XXXIII, pp. 91-131.
- Monitoring and Assessment Report 2006* (2007).** Canada Employment Insurance Commission.
- Nicholson, Walter. (2001)** “The Design of Summative Evaluations for the Employment Benefits Support Measures.” Ottawa. Human Resources Development Canada.
- Schochet, Peter Z. and Burghardt, John. (2007)** “Using Propensity Scoring to Estimate Program-Related Subgroup Impacts in Experimental Program Evaluations.” Draft, Mathematica Policy Research
- Szabo, Les. (2007)** “Report on Validation of the Findings of the Canada-Alberta LMDA Summative Evaluation.” Ottawa, HRSDC.
- Vroman, Wayne. (1997)** *Self-Employment Assistance: Revised Report*. Washington, DC. The Urban Institute.
- Waslander, Bert. (2007)** “Technical Report of Quantitative Analysis in support of the Summative Evaluation of Provincial Benefits and Measures in New Brunswick.” Informetrica, Ltd.

Appendix 1

Use of Statistical Weights in the EBSM Evaluations

All of the EBSM evaluations used statistical weights for at least a portion of their analyses. These procedures were intended to compensate for high rates of non-response to the study surveys and to adjust for stratified sample designs when they were used¹. Because relatively few details are provided about the characteristics of these weighting schemes, it is difficult to know precisely what effect they may have had on the results obtained. But the subject seems an important one to examine, if only to provide some suggestions about how the weighting issue should be handled in future evaluations. With that goal in mind, this brief appendix first describes the general rationale for the use of statistical weights. Next it reviews what information there is about weights from the published EBSM reports. The appendix then concludes with a few general lessons about weighting in evaluations.

1. Rationale for Weighting

If evaluations were conducted using purely random samples of program participants there would be no need to address the issue of weighting. The sample would replicate the population of interest and would therefore be “self-weighting”. There are, however, a variety of reasons why the samples analyzed in an evaluation may not represent simple random samples of the population of interest:

1. Sample designs may call for may over-sampling interventions with relatively few participants;
2. Sample designs may call for other departures from randomness in order to reflect regional or demographic differences;
3. Researchers may not be able to contact individuals who nominally fall into the sampling frame;
4. Other types of non-response may occur, especially for the survey data;
5. Missing data or the unwillingness of respondents to permit linkage of survey and administrative data may reduce samples used in analysis.

All of these issues arose in the EBSM evaluations and all of the research teams adopted weighting techniques for dealing with them. Adoption of these techniques was motivated by two goals. First was the desire to have descriptive statistics from the analysis samples (on, say, basic demographic characteristics) approximate those from the population of interest (say, all EBSM participants). In the absence of weighting sample

¹ A few evaluations also used weighting of the comparison group as their primary matching methodology (Nova Scotia, Prince Edward Island). Assessing the validity of this approach is one aspect of the need to assess all of the matching methods used in the evaluations and has been discussed previously in the body of this report. Because the evaluations that used this method were based primarily on administrative data, weighting issues associated with sample attrition are not as crucial as in those evaluations that primarily used survey data

characteristics could be quite different from those of the underlying population and it was thought that such differences could undermine the perceived validity of the evaluations.

A second and more complex reason for the adoption of weighting was the fear that the impact results may otherwise have been biased by sample attrition. The nature of this concern can best be illustrated with a simple statistical model. Let y represent some post-program outcome of interest (i.e. earnings), x represent a vector of (exogenous) determinants of y , and t represent a randomly assigned (next we take up the case of non-experimental designs) binary treatment indicator. The model to be estimated is:

$$y_i = \beta x_i + \gamma t_i + u_i \quad [1]$$

Where the estimate of γ is to be taken as the treatment impact. If the sample over which equation [1] is to be estimated is purely random, we can safely assume that t_i and u_i are uncorrelated and therefore that standard estimation methods such as ordinary least squares will yield unbiased impact estimates. However, with non-random sample attrition, it is possible that this assumption may fail. If certain types of sample group members are more likely to attrite than others, it is possible that this may create a correlation between t and u that will lead to biased treatment impact estimates. If weights can be developed to compensate for such non-random attrition, such biases can be mitigated (Wooldridge, 2002).

The rationale for weighting in when treatments are not randomly assigned is much the same – that is, the intent of weighting is to “reverse” any potential biases in the estimation of treatment impacts that may have arisen from non-random attrition. The complicating factor in this case is that some of the variables that predict attrition may also predict participation. In this case, impact estimates may be sensitive to precisely how weights are estimated and whether or not they are used as part of the matching process (for example, in the propensity score equation). In general, if weights can be correctly estimated, they should be used in all aspects of the matching process. Econometric literature on this issue is sparse, however. So the best approach in evaluations may be for the authors to provide estimates under several different weighting scenarios.

2. Weighting in the Evaluations

Few of the EBSM summative evaluations provide details on their weighting procedures. Perhaps the most explicit coverage is in Annex G of the British Columbia evaluation. There the authors describe how they estimated a survey response equation based on EI administrative data on a few variables such as gender, age, and EI receipt. This equation allowed them to assign a predicted response rate to each survey respondent and the reciprocal of these rates were used as analysis weights. After normalization, such weights ranged from 0.05 to 84. Clearly such a wide range would appear to have the potential for affecting impact estimates substantially, though no un-weighted estimates were presented in the report.

Special data runs from the Newfoundland evaluation provided additional insights on how sample characteristics may have been affected by weighting. For example, as in most surveys, younger workers were found to be much less likely to respond to the Newfoundland survey. Weighting the sample to take this into account reduced the median age by more than three years (from 36.6 years to 33.1 years). The evaluation also found that claimants who exhausted their EI entitlements were significantly more likely to respond to the survey than those who did not. Weighting reduced the exhaustion rate in the survey sample by nearly four percentage points (from 30.6 percent to 26.7 percent). The relationship, if any, between such adjustments and the impact estimates obtained was not studied in detail, however.

3. Conclusions

The issue of sampling weights may become less important as future evaluations increasingly utilize administrative data where non-response is far less serious. In cases where relevant outcomes must be measured by surveys or when evaluations use complex sampling designs, however, the need to use weights will be unavoidable. Because the relationship between such weighting and various non-experimental estimation procedures (i.e. kernel matching) is not well-understood it might be useful to undertake some empirical research on the topic. Some of the data sets from the EBSM evaluations would be ideal for that purpose.

Appendix 2

Social Assistance Outcomes

All of the EBSM evaluations reported estimated impacts on receipt of social assistance (SA) benefits. These were not reported in the body of this report for two reasons: (1) The report sought to follow international practice by focusing only on direct labour market outcomes; and (2) There is somewhat greater uncertainty about the validity of the matching methods used in the evaluations for SA reciprocity. In this appendix we first describe this second reason in somewhat greater detail and then turn to an examination of the SA impacts that were actually obtained.

1. Problems in matching SA recipients

The matching methods used in the EBSM evaluations relied mainly on data from the EI Status Vector file. This file does not provide information about SA receipt. Hence, only a few evaluations were able to use such data in their matching process. Because SA recipients differ along several dimensions from EI recipients generally, it is likely that the samples of SA recipients in the evaluations are atypical of SA recipients generally. They are probably different from the general population of EI recipients also and their participation in EBSM interventions may be governed by different factors than is the case for most participants. It is possible therefore that matching routines will not work very well for SA recipients and that estimated impacts may be biased.

Problems in matching SA recipients may be especially severe for active claimants. In such cases all EBSM participants have had recent labour market experience, so the subsample of SA recipients may be especially skewed toward unusual circumstances. Whether post-program changes in SA receipt can be attributed to participation or to some other unmeasured factor seems especially problematic for this group. The issue may be a bit less important for former claimants since all most such individuals have had a period of limited labour market experience during which SA receipt would be more common.

2. SA results for former claimants

For these reasons we focus here only of SA outcomes for former claimants. Table A.1 provides a summary of such these results. Overall, modest reductions in annual SA benefits for former claimants appear to have been quite common in the evaluations. That was especially true for Targeted Wage Subsidies for which five or the six evaluations with results found significant negative impacts. With the exception of the rather large decline in British Columbia, the other figures were in the \$100-\$270 range. That finding is consistent with the overall evaluation results for TWS where former claimants were found to be especially likely to experience gains in employment and earnings, perhaps because the intervention provided a good path to return to the labour market. Under that interpretation, a reduction in SA benefit receipt would be a reasonable expectation. A more explicit focus on matching former SA recipients would probably be a good addition to future evaluations, however.

Appendix 3

Impacts on Labour Markets and Communities

To Come