The Canadian Journal of Program Evaluation Vol. 27 No. 2 Pages 21–38
ISSN 0834-1516 Copyright © 2013 Canadian Evaluation Society

ASSESSING THE VALUE FOR MONEY OF ACTIVE LABOUR MARKET PROGRAMMING FOR PERSONS WITH DISABILITIES

Greg Mason
PRA Inc.
Department of Economics, University of Manitoba
Winnipeg, Manitoba

Andrew Buchel
PRA Inc.
Winnipeg, Manitoba

Tony Haddad
Dwane UnRuh
Human Resources and Skills Development Canada
Ottawa, Ontario

Abstract: Treasury Board of Canada's new policy on evaluation and its accompanying directive have placed increased pressure on those conducting federal evaluations to not only quantify the impacts of programming but also make measurable assessments of their value. However, making accurate statements about the value for money of programming can be difficult during evaluations. A number of technical and practical challenges can make common approaches infeasible. This article discusses a recent assessment of the value for money undertaken during the evaluation of the Canada-Manitoba Labour Market Agreement for Persons with Disabilities. It demonstrates a number of approaches that can be used to overcome some of the most common barriers to the assessment of value for money in evaluations.

Résumé : La nouvelle politique sur l'évaluation du Conseil du Trésor du Canada et la directive qui l'accompagne placent une pression accrue sur les épaules de ceux qui mènent des évaluations de programmes fédéraux afin de non seulement quantifier l'impact de la programmation, mais aussi d'évaluer sa valeur de façon mesurable. Cependant, il peut s'avérer difficile en cours d'évaluation d'exprimer de façon précise le niveau d'optimisation des ressources (la valeur) des programmes. Un certain nombre de défis techniques et pratiques peuvent nous empêcher d'utiliser des approches conventionnelles. Cet article porte sur l'évalua-

Corresponding author: Greg Mason, PRA Inc., 500-363 Broadway, Winnipeg, MB, Canada, R3C 3N9; <gmason@pra.ca>

tion de l'optimisation des ressources dans le cadre de l'évaluation de l'Entente Canada-Manitoba sur le marché du travail visant les personnes handicapées. Il illustre plusieurs approches qui peuvent permettre de surmonter les obstacles les plus communs à l'évaluation de l'optimisation des ressources dans le cadre d'évaluations de programmes.

INTRODUCTION

In April 2009, the Treasury Board (TB) of Canada's new Policy on Evaluation took effect. The policy was meant to support the provision of objective evidence regarding the performance of government programming. In particular, it notes that evaluations are to provide a "neutral assessment of the value for money … of federal government programs" (Treasury Board Secretariat [TBS], 2009b). The TB *Directive on the Evaluation Function* provides further detail about the assessment of value for money, noting that the demonstration of efficiency and economy should be among the core issues examined during evaluations. Specifically, the directive notes that evaluation should include an "assessment of resource utilization in relation to the production of outputs and progress towards expected outcomes" (TBS, 2009a).

The policy and its accompanying directive have placed increased pressure on those conducting federal evaluations to not only quantify the impacts of programming, but also make measurable assessments of their value. As most experienced evaluators will know, even when program data are readily available, neither of these two activities is simple. Attributing quantifiable outcomes to programming—rather than contextual factors—is difficult. Even when this attribution takes place, making an assessment of the value for money provided by a program may be complicated. Often, evaluations fall back on less rigorous approaches to assessing value for money in the absence of appropriate quantitative analytical techniques.

With that said, some federal departments have developed novel quantitative approaches to assessing value for money. Since well before the TB Policy on Evaluation refocused attention on this issue, Human Resources and Skills Development Canada (HRSDC) has been supporting innovative evaluation work in this area. One recent example involved HRSDC's evaluation of the Canada-Manitoba Labour Market Agreement for Persons with Disabilities (CM-LMAPD). A cost-sharing agreement between the federal government and the

Province of Manitoba, the CM-LMAPD supports active labour market interventions for people with disabilities who were seeking work.

The evaluation of the CM-LMAPD combined a limited treatment analysis of program impacts with a blend of cost imputation and an innovative approach to cost-effectiveness analysis (CEA). By doing so, it lent critical insight into the value for money provided by the interventions cost-shared under the agreement. This was possible despite a considerable number of analytical barriers facing the evaluation. The sections that follow discuss the principles on which the work was built and the details of the analysis. The work serves as an example of how value for money may be assessed quantitatively in even difficult evaluation settings.

IMPACT ASSESSMENT IN ACTIVE LABOUR MARKET PROGRAMMING

Active labour market programs (ALMPs) have a long history in Canada. With their focus on re-employment, they differ fundamentally from passive labour market programming meant to support individuals' incomes while unemployed (World Bank, 2011). The Employment Insurance (EI) program includes examples of both cash payments to the unemployed through EI Part I programming, and re-employment interventions through EI Part II.

Despite the range of ALMPs offered in Canada both federally and provincially, most fall into one of three categories. These include interventions meant to

- increase the quality of the labour supply—education, training, etc.
- increase labour demand—public work projects, targeted wage subsidies, etc.
- improve the match between job seekers and employers—job search assistance, labour market information, etc. (World Bank, 2011)

The ALMPs cost-shared under the CM-LMAPD fall principally into the first and third categories.

As Mason and Tereraho (2007) note, assessing ALMPs normally requires two main steps. First, measured outcomes must be linked causally to programming. That is, one must show that the activi-

ties and outputs produced through the programming contribute to expected outcomes (2007, p. 3). For most quantitative assessments, it is also critical that the contribution made by the program be measurable and well defined. For example, in the case of an ALMP like a training program, one would be interested in the marginal impact of the program on participant employment—not simply the fact that the two are causally linked.

Randomized experimentation continues to be the "gold standard" for causal inference in most research settings. For example, as Imbens and Wooldridge (2008) note, the medical literature has a long history of experimentation—at times arguing that it is the only reliable means of causal inference. The authors point to the United States Food and Drug Administration's requirement for randomized experimental evidence when approving drugs and medical procedures as but one example of the approach's prominence in these research circles (2008, p. 15).

However, they also note that randomized experimentation has not had the same level of prominence in economics (Imbens & Wooldridge, 2008, p. 15). Although authors such as Greenberg, Shroder, and Onstott point to economic experimentation as early as 1962 (1999, p. 158), Imbens and Wooldridge stress that many other forms of analysis have been used in economics to explore causality (2008, p. 15). The same is true of the social sciences—evaluation in particular—where a broad range of methods are more commonly used.

Notwithstanding important criticisms of nonexperimental methods, such as those of LaLonde (1986), Imbens and Wooldridge point to a number of practical reasons for moving beyond randomized experimentation in these research settings. They note that many treatments of interest—including ALMPs—are readily apparent to individuals both receiving and delivering the interventions. This makes it impossible to conduct the type of blind or double-blind experiments common to medical research (Imbens & Wooldridge, 2008, pp. 15–16).

More importantly, however, there are also a number of important barriers to random assignment—the key to experimental inference. It is often logistically impossible to establish an experimental design prior to initiating treatment. This is true of most ALMPs that are implemented well before the initiation of any associated evaluation work. In addition, ethical considerations may limit the ability to withhold programming from eligible individuals in order to create

separate program and control groups. In these situations, the evaluator is left with a universal program where all eligible individuals receive some form of treatment.

Quasi-experimentation is therefore used regularly as an alternative to randomized experimentation. Generally, it represents a collection of techniques meant to approximate the conditions in a randomized experiment, thereby allowing for similar causal inferences. Under the assumption that no unobserved factors exist that are associated with both treatment assignment and outcomes, a number of techniques may be used successfully. These may include various regression techniques with multiple controls or matching approaches that assign observations into treatment and comparison groups (Imbens & Wooldridge, 2008, pp. 19, 31).

Even when the assumption discussed above does not hold, additional techniques may be used to minimize biases. For example, Difference-in-Differences (DID) methods can control for permanent differences between treated and untreated individuals (Imbens & Wooldridge, 2008, pp. 64–65) even when these are unmeasured. This provides a number of possibilities for quantifying the attributable impact of programming, and meeting the first requirement of assessing value for money. However, as mentioned above, there remains a second requirement: namely, the costs of programming must be accurately defined (Mason & Tereraho, 2007, p. 3).

Unfortunately, accurately measuring program costs can be difficult for most evaluations. Part of this stems from the variety of ways that one can view a program. As Greenberg and Appenzeller (1998) note, program costs may be seen from a number of different perspectives. For example, government may define program costs to include their own expenditures on staff salaries and direct costs. However, from the perspective of the participant, additional personal expenditures during their participation may merit consideration (1998, p. 8).

Even when the perspective is well understood, it is easy to neglect important costs. Greenberg and Appenzeller (1998) state that in programming such as ALMPs, staff salaries, benefits, purchased goods, office space, and vendor payments should all be included. However, they also point to less readily measured items. These include, for example, subsidies to employers to increase labour demand (1998, p. 10). Without considering these, the cost of programming will be understated.

Once the attributable impacts of programming are quantified and program costs are well understood, the assessment of value for money still requires some way of relating these two items to provide an indication of program efficiency. Normally, one of two techniques is used. The first involves cost-benefit analysis (CBA), while the second involves cost-effectiveness analysis (CEA).[1] As Garber and Phelps (1997) note—citing Phelps and Mushlin (1991)—these two techniques are normally treated separately. In the case of CBA, the cost of programming is related to its monetized benefits, typically using a ratio. In the case of CEA, this ratio relates the cost of programming to its outcomes in their natural units (Garber & Phelps, 1997, p. 2).

The fact that CEA presents outcomes in their natural units greatly simplifies the value for money analysis. Using outcomes in their natural units means there is no need to monetize the benefits associated with program outcomes—often a challenging activity requiring judgements about whose valuation is more appropriate in the context of the work. For example, the value of an education as a result of an ALMP may be very different for the delivery organization than for the client. As a result, CEA has become common in evaluation work.

Strictly speaking, CEA only requires a single ratio relating costs to outcomes. This allows the analysis to make statements about the program's per-unit cost of producing specific results. However, without a point of comparison, these ratios often lose meaning. More often, the CEA ratio of a program is compared to alternatives in order to provide more insight. This type of comparison allows one to answer more complicated questions such as those suggested by Greenberg, Michalopoulos, and Robins, and relate these back to the costs of delivering various forms of programming (Greenberg, Michalopoulos, & Robins, 2003, p. 32). This is where more meaningful statements of value for money are possible.

ASSESSING THE CEA OF THE CM-LMAPD

As discussed above, the CM-LMAPD is a cost-sharing initiative between the federal government and the Province of Manitoba that is meant to provide vocational support for people with disabilities.[2] The principal program examined in the evaluation—The Vocational Rehabilitation (VR) Program—was an extension of previous provincial initiatives, making a clear start and end date for the program difficult to establish. More importantly in the context of the current discussion, the program was unique in the province, making a point of comparison for its CEA impossible to find. In addition, the full

cost of programming was "buried" in the operational costs of the provincial department operating VR. Here, the analysis relied on an innovative cost projection approach along with a limited treatment approach to CEA.

Most employment programs for economically disadvantaged clients (e.g., social assistance, employment insurance) rely on a process of client assessment followed by the provision of a range of labour market interventions. These interventions often fall into discrete categories, including basic education, upgrading, skills enhancement, job search, wage subsidies, and others. Within each of these categories, it is possible to have variation in the nature of individual interventions related to their intensity, frequency, and duration.

Assessment of the success of these programs often centres on measuring the increase in "employability" among participants. This may involve determining the extent to which participants actually work in paid occupations following their programming. Less directly, precursors to employment such as improvements in education or the achievement of other training goals can point to increased employability. Even less direct measures, such as a decreased reliance on social assistance or employment insurance benefits, are also often used in an attempt to assess program success.

Outcomes of these types are normally measured in very straightforward ways. This involves the use of self-reported employment or education levels, earnings and benefit information from the tax system, or records from administrative datasets such as those used in social assistance and employment insurance programs. While gathering this information may be difficult in some evaluations, it is normally the attribution of changes in these measures to the program under study that is most challenging. Typically, this requires some comparison of outcomes for participants to those that would have occurred in the absence of the program or under some alternative. Most quasi-experimental evaluations allow for this by identifying a non-participant group whose outcomes provide this point of comparison.

The VR program, which was cost-shared under the CM-LMAPD, presented two challenges to CEA during its evaluation. First, because it was a universal program, all eligible clients were necessarily enrolled and a conventional quasi-experimental design was not feasible. This was because there were no similar individuals left in the province from which to construct the type of non-participant group discussed above. Second, the cost information available for the analysis was limited to direct training costs and excluded the salaries of the voca-

tional counsellors charged with delivering the program. Using only these limited costs would have understated the resource requirements of the program by a significant margin, and created a positive bias in the program's assessed efficiency.

To address the lack of a non-participant comparison group, the evaluation of the CM-LMAPD exploited a feature common to many labour market programs—namely, that participants will undertake different combinations of interventions within fixed categories. In the case of the VR program, all participants received ongoing counselling, but then received a personalized combination of other interventions such as training, education, and job placement. In this context, the counselling represented a baseline, minimal intervention among all of those who participated in VR.

This baseline treatment allowed the evaluation to use the pool of program participants to create a self-referencing comparison group. Here, the outcomes for those receiving the lowest level of intervention—namely, counselling—were compared to those receiving other forms of intervention. Rather than determining the overall impact of the program relative to no programming, this allowed the evaluation to assess the marginal impact of various degrees of additional participation beyond counselling for certain types of interventions. This approach is normally referred to as a "dose response" or "limited treatment analysis." It is analogous to testing the incremental impact of various drug dosages on pain regulation relative to some minimal dose.

For the CM-LMAPD, two important types of marginal impacts were calculated. First, a marginal impact for participation in a specific type of intervention—for example, training—was calculated. One would expect this impact for anyone who had taken one or more training interventions during the program period. Second, a marginal impact for each iteration of a specific form of intervention was calculated. Thus, if a client participated in three training interventions, one would expect three times this estimated impact for this client. These two forms of marginal impact, across the six forms of intervention identified for the VR program, provided considerable flexibility to define individual program impacts.

Table 1 presents estimated marginal impacts on the change in clients' nominal weekly earnings between 2003 and 2007. The two types of marginal impacts discussed immediately above are noted at the bottom of the table. Other impacts related to client characteristics and other aspects of program delivery are noted throughout.

Readers will note that many of the estimated impacts identified in the table align closely with intuitive expectations. For example, employment interventions are associated with a positive change in weekly earnings.[3]

Table 1
Regression Results: 2003–2007 Change in Weekly Earnings (mean = $39.36)

Variable	*Type*	*Mean*	*Coefficient*	*P-value*
Constant	constant	n/a	**231.55**	**0.00**
(Male gender)	M.E. dummy	0.54	-	-
Female gender	M.E. dummy	0.46	3.47	0.84
(Non-Aboriginal)	M.E. dummy	0.95	-	-
Aboriginal	M.E. dummy	0.05	**-87.52**	**0.03**
(Non-minority)	M.E. dummy	0.02	-	-
Minority	M.E. dummy	0.98	80.47	0.08
(Single)	M.E. dummy	0.17	-	-
Married	M.E. dummy	0.72	**-48.95**	**0.02**
Previously married	M.E. dummy	0.11	-43.99	0.41
Number of children under 7 in 2003	count	0.41	-33.39	0.42
Cognitive disability	N.E. dummy	0.36	**-29.76**	**0.09**
Physical disability	N.E. dummy	0.50	2.97	0.89
Psychiatric disability	N.E. dummy	0.38	9.33	0.77
Hearing disability	N.E. dummy	0.09	-56.51	0.07
Vision disability	N.E. dummy	0.17	-90.28	0.15
Learning disability	N.E. dummy	0.33	**41.37**	**0.03**
Other disability	N.E. dummy	0.04	21.35	0.66
Disability onset age	years	22.19	-2.57	0.00
Self-reported disability severity	5-point scale	2.75	**-45.56**	**0.00**
(Urban service delivery region)	M.E. dummy	0.72	-	-
Rural service delivery region	M.E. dummy	0.28	-16.37	0.37
(Provincial service delivery)	M.E. dummy	0.56	-	-
Society for Manitobans with Disabilities service delivery	M.E. dummy	0.25	**57.61**	**0.09**
Canadian Paraplegic Association service delivery	M.E. dummy	0.04	130.76	0.37
Canadian National Institute for the Blind service delivery	M.E. dummy	0.14	**151.23**	**0.04**
Other service delivery	M.E. dummy	0.01	187.26	0.21
Number of assessment interventions during the PP	count	0.69	**14.97**	**0.03**
Number of training interventions during the PP	count	3.47	0.85	0.37
Number of education interventions during the PP	count	4.75	-0.63	0.52
Number of upgrading interventions during the PP	count	0.48	1.41	0.52
Number of employment interventions during the PP	count	0.27	1.52	0.80

(continued next page)

Variable	*Type*	*Mean*	*Coefficient*	*P-value*
Number of follow-up interventions during the PP	count	1.15	**-4.23**	**0.07**
Presence of an assessment interventions during the PP	N.E. dummy	0.21	-30.10	0.30
Presence of a training interventions during the PP	N.E. dummy	0.22	-25.26	0.32
Presence of an education interventions during the PP	N.E. dummy	0.24	41.77	0.15
Presence of an upgrading interventions during the PP	N.E. dummy	0.04	-0.62	0.99
Presence of an employment interventions during the PP	N.E. dummy	0.08	**60.15**	**0.10**
Presence of a follow-up interventions during the PP	N.E. dummy	0.10	75.54	0.06
Note: M.E.—mutually exclusive; N.E.—non-mutually exclusive; PP—program period			**Sample: 1062** **Adj. R^2: 0.0515**	

Having established the marginal impacts of various program interventions relative to the baseline counselling activity, we next needed to address the limited cost information. The administrative data available for the evaluation included information on funded or direct costs of programming. For example, it included records of payments to non-governmental service providers who delivered training or other programming, the costs of materials for clients, and the costs of client transportation. However, these data did not include the cost of the counselling activities undertaken by the VR vocational counsellors. Like many program costs, these were embedded in the overheads of a department.

Developing a more accurate estimate of the program's expenditures that incorporated these counselling costs involved four distinct steps. First, 281 currently active VR clients were identified from among the client population. These were specifically selected to include representation across the disability populations served by the program, thereby capturing variation in programming that was disability-specific. Counsellors of these clients were sent a short survey, which included the following question[4]:

> On average, over the entire time that you have been this client's counsellor, how many hours per month did you spend on this client's case? If you have been this client's counsellor for many years, please consider only the time since the start of 2004. We would like you to consider your time spent on any activity related to this case including, but not limited to, assessments, counselling, monitoring, reporting, paperwork, travel, meetings, and discussions with the client.

Second, responses from the 270 completed surveys were combined with demographic and service provider information from the Depart-

ment of Family Services and Housing—the department operating the VR program. We ran an ordinary least squares (OLS) regression of the demographic and service provider characteristics on the monthly hours spent by counsellors on their clients' cases.

The results, as shown in Table 2, provided an indication of the association between each characteristic and the hours counsellors devoted to each case. For example, the constant term's estimate would suggest that counsellors would devote about two hours of time per month, on average, to individuals with baseline characteristics—male, non-Aboriginal, non-minority, and so on. Those with other characteristics would receive more or less time, on average, based on the associated estimates. As one might expect, the coefficients were in line with the expectation that more difficult cases required more counselling time.

Table 2
Regression Results – Counsellor Hours per Month (mean = 2.16)

Variable	*Mean*	*Coefficient*	*P-value*
Constant	n/a	**2.05**	**0.01**
(Male gender)	0.61	-	-
Female gender	0.39	0.10	0.74
(Non-Aboriginal)	0.98	-	-
Aboriginal	0.02	-1.14	0.26
(Non-minority)	0.99	-	-
Minority	0.01	**3.98**	**0.01**
Age	35.09	-0.01	0.36
Cognitive disability	0.17	0.20	0.78
Physical disability	0.30	**5.43**	**0.00**
Psychiatric disability	0.28	1.08	0.10
Hearing disability	0.09	**6.34**	**0.00**
Vision disability	0.13	0.58	0.73
Learning disability	0.14	-0.61	0.35
(Urban service delivery region)	0.69	-	-
Rural service delivery region	0.31	0.17	0.612
(Provincial service delivery)	0.52	-	-
Society for Manitobans with Disabilities service delivery	0.31	**-6.06**	**0.00**
Canadian Paraplegic Association service delivery	0.06	**-5.16**	**0.00**
Canadian National Institute for the Blind service delivery	0.12	0.61	0.74
		Sample: 270	
		Adj. R^2: 0.1508	

The third step involved using the estimated values above to impute counselling time to the full client sample available for the evaluation. Recall that this was the sample from which the marginal intervention impacts discussed above were developed. The calculation involved multiplying the full client sample values for the independent variables noted in Table 2 by the coefficients in the table. This provided an estimated monthly counselling cost for each individual in the sample.

The fourth step in the process averaged these estimated counselling costs over the program period and combined them with the costs of the program's interventions and other direct expenses. As noted above, the evaluation identified six distinct forms of interventions and one additional category of other funded services. The average cost for one instance of each type of intervention and the average cost of the other funded services over the program period were calculated. These are represented in Table 3.

Table 3
Average Cost of Funded Services – Direct Costs (*n* = 1,062)

Funded service	*Mean*	*Standard deviation*
Assessment interventions	$160.29	$387.94
Training interventions	$286.08	$818.21
Education interventions	$951.14	$2,597.26
Upgrading interventions	$39.16	$366.40
Employment interventions	$134.22	$719.61
Follow-up interventions	$83.92	$339.64
All other funded services	$137.72	$451.19

As a result of these four steps, the evaluation now had two forms of marginal impact (relative to the baseline counselling intervention) and the average cost of each type of intervention, along with the average counselling costs and average other funded services costs over the program period. This allowed for a CEA that mapped out the relative cost-effectiveness of successive interventions of a specific type. This answered the question: "Is there increasing or decreasing cost-effectiveness of successive VR interventions relative to baseline counselling?"

The calculation used for this CEA is shown below, where an intervention of a specific type is identified by the term "X."

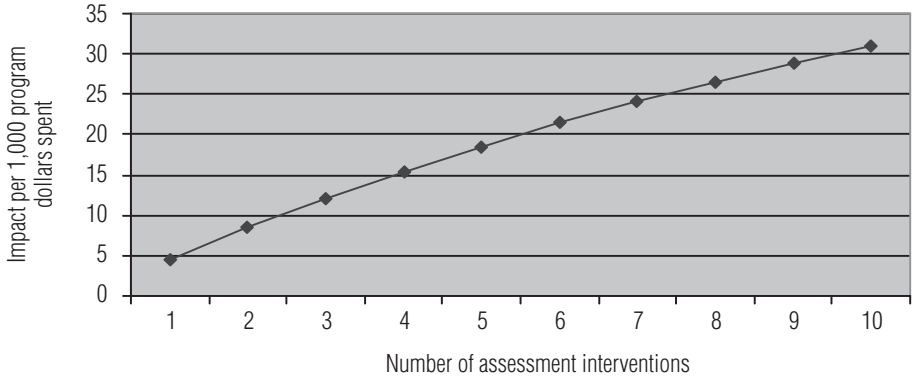
(marginal impact of the presence of X during the program period) +
(marginal impact of one iteration of X during the program period) *
(number of iterations of X during the program period)

(average counselling cost during the program period) +
(average non-intervention-funded service cost during the program period for the group) +
(average cost for one iteration of X) * (number of iterations of X during the program period)

To illustrate how this CEA would operate, consider one outcome used during the evaluation—the change in nominal weekly earnings between 2003 and 2007—and one specific form of intervention—client assessment. The marginal impacts calculated for the assessment intervention suggest that while the presence of an assessment intervention is associated with a reduction in weekly earnings (based on its point estimate), each iteration of the assessment intervention is associated with a slightly smaller, yet positive impact on earnings.

This, combined with the relatively low average cost of each assessment intervention, suggests that subsequent assessment activities have increasing cost-effectiveness relative to the first. In simpler terms, additional assessments—when necessary in the context of VR program delivery—provide good value for money. This may be the result of additional assessment activity more accurately defining client needs and aligning individuals with appropriate interventions. The figure below helps demonstrate this point by showing the average impact of an assessment per $1,000 of program expenditure, when 1 to 10 assessments take place.

Figure 1
Assessment Intervention



DISCUSSION

In the face of current federal requirements, evaluations in Canada will need to carefully consider the value for money provided by program interventions. Like many programs, ALMPs may have a number of characteristics that make assessing value for money difficult. Determining their attributable impact on outcomes, dealing with their universal nature, and fully capturing all costs can all prove challenging during evaluations. However, as the discussion above has shown, creative approaches can still provide insights.

In the case of the CM-LMAPD's VR program, a limited treatment analysis allowed the evaluation to estimate the marginal impact of various program components. A counsellor survey followed by a simple imputation process helped ensure that a fuller accounting of costs took place. Finally, assessing the average cost-effectiveness of interventions based on their frequency allowed for statements about the value for money provided by different levels of program treatment.

While not an assessment of the relative value for money of the program as a whole, this analysis provides important insight into the value of various intensities of programming. This type of information is critical to the improvement of programming and effective use of internal resources. Future evaluations can certainly use and improve on these techniques to explore similar aspects of programming.

LIMITATIONS

From the discussion above, it is possible to see how the analysis of the CM-LMAPD overcame many challenges. However, other challenges, such as possible endogeneity in the impact estimates and uncounted costs, remain. While these problems require examination of the CEA results with some caution, value remains in pursuing this approach.

As experienced evaluators will know, program evaluation in the context of government-funded programming often involve data and design limitations not present in other research settings. Like the CM-LMAPD cost-shared programs, initiatives may be implemented in advance of an established evaluation design. Data collection is not undertaken to support the strongest possible evaluative insights, but rather to support program administration. In these situations,

programs serve clients first, and evaluations must find the most effective way to use the available information to support value for money assessments.

While it is true that an experimental analysis of impacts during the CM-LMAPD evaluation would have provided more reliable inputs into its CEA, the CM-LMAPD was a cost-sharing initiative supporting programming that had existed for decades. The retrospective nature of the evaluation thus precluded this type of design and required choices to be made about how to define the program period studied in the analysis. In addition, the VR program examined in the work was nearly universal, meaning that an experimental analysis would have required withdrawing services from a number of Manitobans with disabilities. Even were a prospective design possible, this would have been unethical, and certainly would not have been politically sanctioned.

Alternatives such as instrumental variables analysis and quasi-experimentation involving data from other provinces—while potentially providing improvements in impact estimates—were not feasible. Like many other program evaluations, the evaluation of the CM-LMAPD lacked effective instruments for the multidimensional intervention measures associated with programming. In addition, because this was an evaluation of a Manitoba-specific cost-sharing agreement, comparable data from other provinces were not available. In fact, some key data elements for this program were not available in Manitoba, thus requiring the cost imputation undertaken as part of the analysis rather than a straightforward apportionment of costs across all provincial programming.

Despite these limitations and the current inability to revisit, redesign, and re-implement the CM-LMAPD, the CEA applied to this program serves an important evaluative purpose. It provides additional quantitative insight within a broader evaluation that used a broad array of mixed methods. With the current Treasury Board focus on quantifiable assessments of value for money, this seems a preferable option to largely abandoning quantitative analysis and relying solely on qualitative methods in the face of design and data challenges.

NOTES

1 Cost-utility analysis (CUA) is sometimes cited as a third alternative, but is less readily considered in most evaluation work.

2 A full version the CM-LMAPD evaluation report is available through HRSDC at http://www.hrsdc.gc.ca/eng/publications_resources/evaluation/2010/sp_949_05_10e/page03.shtml
The report details the qualitative and quantitative research activities undertaken during the evaluation in addition to the CEA.

3 Readers will note that the adjusted R-squared in the table is low, suggesting that the independent variables in the model account for only a portion of the variation in the model's dependent variable.

4 The CM-LMAPD came into force in 2004; as such, this was defined as the start of the program period for the purposes of the evaluation.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the comments of the anonymous referees. Morley Gunderson, Derek Hum, Walter Nicholson and Wayne Simpson also offered much valuable advice throughout this research. Any and all remaining errors remain ours.

REFERENCES

Garber, A. M., & Phelps, C. E. (1997). Economic foundations of cost-effectiveness analysis. *Journal of Health Economics*, *16*(1), 1–31. doi:10.1016/S0167-6296(96)00506-1

Greenberg, D. H., & Appenzeller, U. (1998). *Cost analysis step by step: A how-to guide for planners and providers of welfare-to-work and other employment and training programs* (Report Number 10/1998). Manpower Research Demonstration Corporation. Retrieved from http://www.mdrc.org/cost-analysis-step-step

Greenberg, D., Michalopoulos, C., & Robins, P. (2003). Meta-analysis of government-sponsored training programs, A. *Industrial and Labor Relations Review, 57*, 31.

Greenberg, D., Shroder, M., & Onstott, M. (1999). The social experiment market. *Journal of Economic Perspectives*, *13*(3), 157–172.

Imbens, G. W., & Wooldridge, J. M. (2008). *Recent development in the econometrics of program evaluation* (Working Paper 14251). Cambridge, MA: National Bureau of Economic Research.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, *76*(4), 604–620.

Mason, G., & Tereraho, M. (2007). Value-for-money analysis of active labour market programs. *Canadian Journal of Program Evaluation*, *22*(1), 1–29.

Phelps, C. E., & Mushlin, A. I. (1991). On the (near) equivalence of cost-effectiveness and cost-benefit analyses. *International Journal of Technology Assessment in Health Care*, *7*(01), 12–21. doi:10.1017/S0266462300004803

Treasury Board Secretariat. (2009a). Directive on the evaluation function. Retrieved from http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=15681&section=text

Treasury Board Secretariat. (2009b). Policy on evaluation. Retrieved from http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=15024

World Bank. (2011). Labor markets: Active labor market programs & activation policies. Retrieved from http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALPROTECTION/EXTLM/0,,contentMDK:20223809~menuPK:584846~pagePK:148956~piPK:216618~theSitePK:390615,00.html

Greg Mason is a certified management consultant and economist specializing in social and economic research, program evaluation, and policy analysis for government and private sector clients. Greg teaches research methodology, program evaluation, and value-for-money analysis to professionals, public servants, and university students. He also holds a reduced appointment (tenured) in the Department of Economics, University of Manitoba, and regularly publishes in the academic/technical literature (www.gregorymason.ca).

Andrew Buchel is an economist with extensive training and experience in applied policy and econometrics. His assignments have

featured the processing and analysis of large databases using a wide range of statistical and econometric methods. Andrew also has delivered professional training in program evaluation methods to public servants. At PRA, Inc., he has assumed responsibility for managing increasingly complex studies, involving teams with diverse skills and subject matter expertise.

Tony Haddad is a director within the Evaluation Directorate at Human Resources and Skills Development Canada. He has held this position since 2006, overseeing a team responsible for the evaluation of partnership-based labour market program programs, including the Labour Market Development Agreements, Labour Market Agreements, Youth Employment Strategy, Aboriginal Human Resources Development Strategy, Aboriginal Skills and Partnerships Program, and Targeted Initiative for Older Workers. He has also held various other positions within HRSDC, including policy advisor and evaluation manager. He holds a Ph.D. in sociology from York University.

Dwane UnRuh is an evaluation manager in the Evaluation Directorate of Human Resources and Skills Development Canada. He has been responsible for developing evaluation strategies and overseeing the evaluations of social development programming provided by HRSDC, including programs for youth, older workers, Aboriginals, and persons with disabilities. He also has extensive experience working in the not-for-profit and charitable sectors, in particular managing programs targeting persons with serious mental illness.