Health Canada Santé Canada

Synthesis of Findings of 2005-2006 Health Canada Departmental-Audit-and-Evaluation-Committee-Approved

Evaluation Reports

Presented to

Health Canada
Departmental Audit and Evaluation Committee

March 8, 2007

Canada

Health Canada Santé Canada

Synthesis of Findings of 2005-2006 Health Canada Departmental-Audit-and-Evaluation-Committee - Approved Evaluation Reports

December 2006

Prepared by:
Departmental Performance Measurement and Evaluation Directorate/
Direction de la mesure et de l'évaluation du rendement du Ministère
Chief Financial Officer Branch/
Direction générale du contrôleur ministériel

Canada

Table of Contents

Page

ACRONYMS . . . . . **ii**

INTRODUCTION **1**

METHODS **2**

SYNTHESIS FINDINGS **4**

Implementation Issues 4

Evidence of Programmatic Impact 7

Quality of Evaluation Reports 8

REVIEW OF MANAGEMENT ACTION PLANS **10**

CONCLUSION **11**

Acronyms

USED IN THIS REPORT

ADMs	Assistant Deputy Ministers
BF/BHC	(First Nations and Inuit Health Branch's) Brighter Futures and Building Healthy Communities Programs
CHIPP	Canada Health Infostructure Partnership Program
DAEC	Departmental Audit and Evaluation Committee
DPMED	Departmental Performance Measurement and Evaluation Directorate
FNI	First Nations and Inuit
FNIH	First Nations and Inuit Health
FNIHB	First Nations and Inuit Health Branch
HCSPG & CP	Health Care Strategies and Policy Grant and Contribution Programs
ICT	Information and Communication Technologies
MOU	Memoranda of Understanding
NIHB	Non-Insured Health Benefits
PMRA	Pest Management Regulatory Agency
RDGs	Regional Directors General
RMDD	Research Management and Dissemination Division

INTRODUCTION

Health Canada's on-going evaluation program provides objective and independent information on the performance of Health Canada's programs, policies, initiatives and functions and promotes performance measurement, to increase the Department's effectiveness, efficiency and economy.

In the last three years, Health Canada's evaluation function has produced, on average, some 9 to 11 evaluation studies each year. Most evaluation studies, however, focus on individual policies, programs or initiatives. Also, when they are submitted to the Departmental Audit and Evaluation Committee (DAEC) for approval[1], the reports tend to be reviewed in a "stand-alone" fashion because they are submitted individually -- as soon as they are completed, at different times during the year. Furthermore, when reports are approved by DAEC without formal discussion (such as electronically through e-mail), issues are, often, not raised for a more general consideration.

As a result, lessons learned, innovations and other issues raised in the evaluation of one policy, program or initiative are not always widely shared or looked at for more general application outside the area that was evaluated.

The purpose of this synthesis is to summarize key findings and recommendations of evaluation studies that were approved by DAEC in 2005-06 so as to identify (and, where appropriate, generalize from) lessons learned, innovations, and other information that could help Health Canada improve management, policy or program effectiveness outside the areas evaluated, including in the development and design of new policies and programs. The synthesis will also examine evaluation study design and processes so as to improve the evaluation function's contribution to decision-making in the Department.

To contribute to this goal, two research designs – evaluation synthesis and meta-evaluation – will be used to critically review the 2005-06 evaluation reports.

[1] Health Canada policy is that all evaluations completed in the Department and the management action plans that respond to evaluation findings, conclusions and recommendations must be submitted to the Departmental Audit and Evaluation Committee (DAEC) for review and approval. DAEC also monitors the implementation of the action plans.

An evaluation synthesis is a systematic procedure for organizing findings from several evaluation studies.[2] The evaluator gathers results from different evaluation reports performed by different people at different times and places, and "aggregates the findings from individual studies to provide a conclusion more credible than that of any one study."[1, 3]

A closely related technique is **meta-evaluations**, which are systematic reviews of evaluations (and evaluators) to determine the quality of their processes and findings. [4, 5] A meta-evaluation checks evaluations for problems such as bias, technical errors, administrative and legal hurdles and other unpredictable difficulties. Meta-evaluations enable organizations to ask questions as to the clarity of evaluation findings, the appropriateness of the evaluation design, whether conclusions were justified by the data and the feasibility of study results or recommendations.

This report consists of four main sections. The introduction provides brief insights into the purpose of this review. The methods section describes the methodology used in the review. Following that is a discussion on the review's findings. The final section summarises the findings and conclusions of this review.

METHODS

This review examined eleven evaluations for four attributes:

Reliability is the consistency or stability of a measure, test, tool or technique. It addresses such questions as: how dependable are the evaluation findings, can they be reproduced or are they simply random and, as such, not meaningful? Faulty data collection procedures and sampling errors could lead to an unrepresentative study population or unreliable information, which could result in defective evaluation conclusions.

Validity*:* The focus here is on internal and external validity. *Internal validity* relates to the extent to which the findings are true. Does the evaluation design answer the questions it is intended to answer? Are the results accurate and unequivocal?

---

2 General Accounting Office. The Evaluation Synthesis. GAO/PEMD-10.1.2 March 1992. Washington, DC: GAO, page 6.

3 General Accounting Office. Designing Evaluations. GAO/PEMD-10.1.4 March 1991. Washington, DC: GAO, page 64.

4 Cooksy, L.J. & Caracelli, V. J. (2005). Quality, Context, and Use: Issues in Achieving the Goals of Meta-Evaluation. American Journal of Evaluation, 26, 31-42

5 Henry, G. T. & Mark, M.M. (2003). Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. American Journal of Evaluation, 24, 293-314

External validity refers to the generalizability of evaluation findings. To what populations and programs can the findings be applied? Does the information hold for only the sample from which it was collected or for other groups (thus, to what extent can the evaluation findings be safely applied to other situations)?

Objectivity refers to the notion that findings and conclusions are based on factual evidence obtained through scientific methods or reasoning. Are the results highly dependent on the unique experiences, perceptions and biases of an evaluator or would other evaluators, in general, disagree or agree with the findings? Generally, if different but equally competent experts cannot arrive at the same or similar conclusions about the evaluation findings, then the true essence of the evaluation results would be subject to question.

Relevance*:* This attribute deals with whether the findings respond to the purposes of the evaluation. The evaluation findings need to match the information being sought by the evaluation audience for the results to be used. Gathering data that could be used to appropriately answer the evaluation questions of interest helps in resolving issues of relevance.

EVALUATIONS APPROVED BY DAEC IN 2005-06

In all, 11 evaluation studies were reviewed for this report. Most of the evaluations were formative in nature, but there were a few summative reports and impact assessments. Completed and DAEC-approved evaluations and reviews in fiscal year 2005-06 are as follows:

1. Canada Health Infostructure Partnership Program (CHIPP)
2. Pest Management Regulatory Agency (PMRA) Cost Recovery Initiative
3. First Nations and Inuit Health Branch's Brighter Futures and Building Healthy Communities Programs (BF/BHC)
4. Health Canada Innovation Fund
5. Health Care Strategies and Policy Grant and Contribution Programs (HCSPG & CP)
6. First Nations and Inuit Health Transfer Policy
7. Health Transition Fund
8. Memoranda of Understanding (MOU) between the Assistant Deputy Ministers (ADMs) and Regional Directors General (RDGs)
9. First Nations and Inuit Health Branch's Non-Insured Health Benefits (NIHB) Pilot Projects
10. Primary Health Care Transition Fund
11. Research Management and Dissemination Division (RMDD)

SYNTHESIS FINDINGS

This section of the synthesis highlights the strengths and weaknesses of the programs or initiatives reviewed. It will touch upon:

- implemental issues;
- whether or not there was sufficient evidence to conclude on the program or initiative's impact on health outcomes; and
- observations on the quality of the evaluation work itself.

IMPLEMENTATION ISSUES

Five key program implementation issues were identified:

(1) Program objectives, goals and operational criteria need to be more clearly articulated

The 11 reports noted that, often, the objectives and goals of programs were not clear enough. This resulted in different interpretations of what intended results were, especially in instances when multiple players were involved.

Some examples:

- The MOU agreements between ADMs and RDGs (report number 8) tended to be written at a high, strategic level and did not have formal "procedures" or "guidelines" for clarifying roles, responsibilities, accountabilities and the specifics of implementation at more operational levels.

- The criteria for administering some programs, such as those covered by report number 3, might not be sufficiently clear in guiding community program managers in deciding on where and how funding received could be allocated. Specifically, criteria on the kinds of activities that were eligible for funding under the BF/BHC programs did not provide enough guidance for First Nations and Inuit (FNI) community program managers to, for example, conclude on whether or not communities with no serious solvent abuse problems, could use funds for that component of the program to fund or supplement non-substance abuse programs.

- Objectives and operational criteria are often developed at early stages of a program. Often, modifications might be made at the program delivery level but corresponding adjustments are not formally reflected at the program authority level, resulting in formal policy statements and directives that no longer reflect what is actually being done. This problem is further aggravated when such changes are not documented (on the part of community or Health Canada regional program managers) to Health Canada managers at the branch or national program level and, when necessary, made consistent across the program.

- Roles and responsibilities are often not clearly articulated. Report number 6, for example, concluded that uncertainty in specific roles and responsibilities among some FNI communities, provinces and FNIHB with regard to dealing with public health threats could result, in the event of an influenza pandemic, in the different players not being able to effectively or coherently deal with the outbreak.

- In some cases, different terminology is used to refer to the same concepts. For example, report number 5 noted that the objectives of the Health Care Strategies and Policy Grant and Contribution Programs (HCSPG & CP) were, at times, referred to or described as 'activities' and, at other times, as 'outcomes'. Document reviews of the program also found that the listed objectives were not consistent across several official policy documents.

(2) Delays in disbursement of funding

Timely allocation of funding was a recurring issue in some of the reports. It was particularly evident in time-sensitive initiatives and programs. In the programs covered by reports 4 and 10, for example, by the time funding was released for some projects, there was not enough time remaining in the life of the programs to effectively implement and assess the impact of the projects. The time required to obtain ministerial approval of contribution agreements was often the main cause of the delays.

(3) Timing of evaluations for program impact

Linked to the disbursement of funding is the time required to assess the effectiveness of programs. Time limited initiatives have built-in evaluation requirements. Delays in implementation or project initiation mean that evaluations, if they are to meet Treasury Board imposed deadlines, will take place too soon after program launch to allow good evidence-based conclusions on performance.

Furthermore, some programs and projects require significant time -- sometimes well beyond the period covered by their funding -- before longer-term outcomes can be observed. Such was the case of CHIPP, on which report number 1 concluded that the evaluation, which was conducted toward the end of the program, could not capture the program's longer-term impacts.

Generally, for time-limited programs, funding for them often end before longer-term impacts can be demonstrated or observed. Consequently, the Department does not follow up with evaluations to assess whether or not the expected longer-term outcomes are actually achieved, at the time that they were expected to be realized. The danger created by this situation is that when the achievements of such initiatives are reviewed from the available evidence, the impact of Health Canada's work is seen as "inconclusive" or "not evident".

(4) Performance measurement weak

Almost all the reports concluded that performance information was lacking. In some instances, such as the program covered by report number 10, a performance measurement framework was developed but it was not implemented. In other cases, such as the program in report number 11, the available information was of limited value for evaluating performance as the data had not been systematically collected or categorised. Report number 6 was an example of a program in which communities involved in the initiative might have systematically collected information but had used different methods for gathering and aggregating their data, making them inconsistent and incomparable.

A few programs, such as that covered in report number 5, had performance indicators and measures. Their usefulness was, however, somewhat compromised because most of the indicators focussed on process and inputs, with little direct linkage to the expected outcomes that the programs were expected to deliver.

In the FNI community initiatives that were evaluated in 2005-06, data collection and reporting were geared more toward program activities and financial transactions (i.e., compliance focused) rather than on performance or outcomes. This may be due to Health Canada's need to track and account for the funds provided for community initiatives. Project managers noted that the sheer number of reports that had to be submitted to the Department meant that they spent more time and effort on completing monthly activity reports than on collecting outcome data.

A further complication with respect to performance measurement in FNIH programs is that although the components of the programs and services received from Health Canada may be the same, FNI communities tend to administer them differently. Modifying programs to meet unique needs is understandable. However, in most cases when these modifications occur, the corresponding adjustments to expected outcomes or program goals were not made. Even when these adjustments are made, though, there may still be a tendency to expect the programs to meet the original "policy outcome" objectives, resulting in "inconclusive" findings.

On a more general level (and applying to FNIH and other programs), grant and contribution programs are often regarded by fund recipients as funding sources. Once received, funds from the different funding sources are often commingled and allocated in accordance with "local priorities", making the linking of "performance" to "inputs" and to individual programs difficult, if not impossible.

(5) Information and knowledge sharing with and among stakeholders not sufficiently actively promoted

An evaluation conclusion of report number 1 was that the program could have achieved more success if stakeholders had shared best practices and lessons learned amongst themselves. This issue was particularly evident in time-limited programs with multiple stakeholders. It appeared that linking projects from different applicants and encouraging them to explore partnerships or

network to improve program outcomes were not emphasized as program administration objectives. Political pressure to have programs "operational" as soon as initiatives were announced may have, in most cases, encouraged the consideration of project applications individually at the expense of more global objectives or approaches.

Dissemination strategies for research-type programs, such as that covered by report 11, were generally weak. Most often, the dissemination of research reports and findings was based on passive rather than active strategies – i.e., using a broad rather than a targeted dissemination strategy or counting on researchers themselves to do the primary dissemination.

Some implementation strengths

While the above has noted some key deficiencies in program implementation, the evaluations and reviews also highlighted some strengths.

New tools created: The CHIPP evaluation noted that the program gathered and presented all the outputs of projects and project evaluations as part of a database of tools, products, protocols, standards, best practices and lessons learned. This allowed users easy access to learnings such that they could be shared and be the launching pad for further research, should parties be so inclined, to accelerate the development of pan-Canadian information and communication technologies (ICT).

Inclusive consultation: All-inclusive consultative processes, such as that used in the program covered by report 10, reconfirmed that involving all interested parties would increase stakeholder buy-in. These programs made concerted efforts to involve interested stakeholder groups, including provinces, territories and national non-government organizations, right from the program design stage. In the case of the program covered by report 10, the result was a program implementation design that was consistent with the primary health care renewal goal of health practitioners understanding and providing health care across Canada from a holistic (rather than a practitioner-in-isolation) perspective.

EVIDENCE ON PROGRAMMATIC IMPACT

Impact assessment is the determination of the extent to which a program produced desired changes in a target population. It normally entails estimation or measurement of the effects of a program, project or activity and thus often requires a comparative analysis of the before and after.

In general, the 11 evaluation and review reports found that Health Canada's ability to determine the impact of its programs is limited by a lack of performance data and, in some cases, of performance frameworks. Key findings were:

(1) Analytical frameworks needed in assessing impact

Analytical frameworks (agreed upon criteria for assessing program impact) were generally lacking. A review of report number 4, for example, showed that there was no framework to determine how program impact would be measured. Without such a framework it is virtually impossible to tease out any intended or unintended effects of a program. Success or failure in such instances are subjective and debatable.

(2) Baseline data missing

Baseline[6] data (the starting point for monitoring results) on programs were non-existent. Most evaluations were, therefore, forced to rely on ex-post analyses, which led to mostly qualitative subjective judgements by program staff or recipients. The studies reviewed relied heavily on program activities and outputs to assess impacts rather than on more rigourous or robust approaches, such as pre- and post-test comparisons of baseline information.

(3) Outcome data weak

Program performance data were generally missing, limited or weak in all the studies reviewed. The main reason for this was programs not placing much emphasis on collecting outcome data. This issue was further compounded by the lack of standardised approaches to measuring impacts. For example, the projects assessed under report number 9 were not comparable because the pilots were administered differently from one another, yielding outcomes that could not be generalized to other FNI communities.

QUALITY OF EVALUATION REPORTS

Program evaluation is the periodic, systematic investigation of – leading to conclusions on – the merit, worth or value of a policy, program or initiative. While evaluation is often considered, to some extent, an "art, there is a core body of knowledge and generally accepted approaches. This synthesis did not systematically examine the actual processes and methodology used in the 11 studies reviewed. It, rather, just reviewed the reports themselves, as written, to determine potential areas for improvement.

Four key issues were identified for comment.

---

[6] Baseline data is the initial information on a program or program components collected prior to the receipt of services or participation in the program. Baseline data are often gathered through intake interviews and observations and are used later for comparing measures that determine changes in a program.

(1) Sampling and data quality could be improved

Methodological rigour is key to useful evaluations and defensible conclusions ... and the ability to generalize from the specific to the general improves information usefulness. For the results of any study to be generalizable, however, the sample has to be representative of the sample frame (or the population in scope for the study).

The evaluation reports reviewed did not demonstrate that the samples used in the studies were representative. Accordingly, the evaluations cannot conclude, within an identifiable or determinable margin of error, the incidence or prevalence of findings (such as outcomes) at the policy, program or initiative level. Also, the possibility of extrapolating lessons from such evaluation studies to similar initiatives becomes impossible.

Also, the quality of data used in the evaluation reports reviewed was generally poor. Their accuracy, consistency, relevance and completeness were, often, questionable. Likewise, none of the programs evaluated, had verifiable data standards to ensure that the data being gathered was bias free.

(2) Objectivity (or "neutrality")

Evaluation studies are supposed to be ("hard") evidence-based, "neutral" assessments of a policy, program or initiative's merit, worth or value, or management processes. The reliability and accuracy of findings and conclusions would be enhanced if they were confirmed by data arising from a variety of approaches or "lines of evidence".

Some of the evaluations reviewed used multiple lines of evidence, such as triangulation[7] in their research. Despite this, it was evident that some "independent" sources were affiliated in diverse ways to the program or subject area been evaluated and evaluation conclusions were often mainly based on subjective opinions. Furthermore, different lines of evidence (e.g., expert panels, literature reviews, key stakeholder interviews, document reviews) could be drawing from the same evidence-base, such as expert-stakeholder interviewees[8] writing the literature or documents reviewed and serving as officers of stakeholder organizations interviewed. The objectivity of evaluation findings based on such interviews, such as in report number 6, can therefore be challenged.

---

[7] Triangulation, is a method of establishing the accuracy of information by comparing three or more types of independent points of view on data sources (for example, interviews, observation, and documentation at different times) and their ability to arrive at the same or similar conclusion.

[8] As used here, the term "stakeholder", is intended to include Health Canada .

(3) Research methods

Evaluation reports should include key methodological decisions or approaches that would allow readers to assess the validity and accuracy of findings and conclusions. Inclusion and exclusion criteria for lines of evidence used would, for example, be important.

The inclusion and exclusion criteria used in recruiting study participants for surveys or interviews were generally missing from the reports. Moreover, very little or no justification was provided to clarify why certain people were excluded or included in surveys or interviews. Clearly outlining these parameters could allow for the easy replication of samples and test the rigour of findings and conclusions by, for example, third parties who wish to do so.

(4) Reports

Evaluation reports must be internally consistent; specifically, evaluation recommendations must be based on and logically flow from evidence-based findings and conclusions articulated in the report.

With the exception of report numbers 4, 5, 8 and 11, evaluation findings did not necessarily match conclusions in some reports. Causality inferences described in some reports lacked the empirical evidence to support their claims. An example was report number 9 that resulted in 41 recommendations without supporting evidence. The Departmental Performance Measurement and Evaluation Directorate's (DPMED) *Evaluation Report Assessment Guide* was used as the yardstick for arriving at the above conclusion.

REVIEW OF MANAGEMENT ACTION PLANS

Health Canada policy is that each evaluation report must include a management action plan. This plan lays out actions and processes that line-management has committed to implement to address the issues identified in the evaluation report. (It also allows line-managers to present their perspective, should it be different from that of the evaluator.) As part of the evaluation report approval process, DPMED assesses the proposed management action plan against the evaluation recommendations (and, where appropriate, against the evaluation conclusions and findings) for reasonableness and the branch head formally commits to implementing the plan prior to its submission to DAEC for approval.

In general, the management action plans do directly and adequately address issues raised in the evaluation. Often, however, stakeholders need to be consulted on the details of specific actions, particularly in First Nations and Inuit health programs. This is recognized in the plans submitted to DAEC and DPMED follows up with program areas as part of its implementation monitoring responsibilities.

Two issues arise with respect to the action plans:

(1) Evaluation results and management action plans "siloed"

One of the benefits of evaluation is the application or adaption of lessons learned to areas not directly examined by an evaluation. The action plans, however, have tended to be developed, more or less exclusively, by and focused on the policy, program or initiative evaluated. Consequently, management actions to address evaluation findings such as the clear articulation of program objectives, goals and operational criteria, and weak performance information do not appear to be taken up (or seen as relevant) in areas and organizations that were not directly implicated in an evaluation.

Indeed, in some cases, evaluation results for a time-limited program that is replaced by a slightly different successor program (and, perhaps named differently) often do not seem to be considered in the development or implementation of the latter.

Also, despite evaluations that have consistently found that information to support good evaluations are often lacking, particularly in the early years of new programs, many Treasury Board submissions are still committing the Department to conducting multiple evaluations over the typical 5-year life-cycle of time-limited or grant and contribution programs.

(2) "Good intentions" for "discontinued" programs

With increasing focus on results and recognizing the time needed for the results of some Health Canada programs to manifest themselves for meaningful measurement, some managers of discontinued programs (such as CHIPP) have committed the Department to actions that are to take place after the program (and the funding) has expired.

While the intentions are good – i.e., in the CHIPP example, to assess the impact of departmental activities at a time when it can be measured so as to conclude on possible policy effectiveness and, thereby, fulfil Health Canada's policy accountability – the probability of actual follow through is apt to be slight because the organization that would be responsible – or the funding – for the action might no longer exist when the work needs to be done.

CONCLUSION

While the need for evaluations and their timely reporting is acknowledged, more often than not, the prerequisites for good evaluations, particularly to assess effectiveness and value-for-money, have been missing. Based on the evaluation reports examined for this review, to improve program effectiveness in terms of outcomes and management, and to improve the quality and useful of the evaluation function itself, Health Canada should consider the following:

- making program objectives, goals and operational criteria clearer and, more particularly, requiring the development of robust management and analytical frameworks that would provide the basis for clear roles and responsibilities and for determining and assessing performance in terms of programmatic or health outcomes and management (process) effectiveness;

- developing and implementing performance measurement initiatives – e.g., performance frameworks, focus on outcomes in addition to that on activities and output, data collection and analysis – to build a strong evidence base to assess policy, program and management effectiveness;

- improving the quality of evaluations by strengthening their methodological rigour and objectivity; and

- using evaluation information as a "strategic" resource – e.g., assessing evaluation findings, conclusions and recommendations for application across the Department (and not just for the policy, program or initiative that was evaluated).