Santé Canada Health Canada

Synthèse des résultats des rapports d'évaluation pour 2005-2006 approuvés par le Comité ministériel de vérification et d'évaluation de Santé Canada

Présenté à

Santé Canada
Comité ministériel de vérification et d`évaluation

8 mars 2007

Canada

Health Canada Santé Canada

Synthèse des résultats des rapports d'évaluation pour 2005-2006 approuvés par le Comité ministériel de vérification et d'évaluation de Santé Canada

Décembre 2006

Préparé par :
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 des matières

Page

INTRODUCTION **1**

MÉTHODES **2**

SYNTHÈSE DES CONSTATATIONS **4**

Problèmes de mise en oeuvre 4

Preuves de l'incidence des programmes 8

Qualité des rapports d'évaluation 10

EXAMEN DES PLANS D'ACTION DE GESTION **12**

CONCLUSION **13**

INTRODUCTION

Le programme d'évaluation continue qu'applique Santé Canada lui permet d'obtenir des informations objectives et indépendantes sur le rendement de ses programmes, politiques, initiatives et fonctions. Il favorise également la mesure de ce rendement dans le but d'accroître l'efficacité, l'efficience et l'économie du Ministère.

Au cours des trois dernières années, la fonction d'évaluation de Santé Canada a produit en moyenne entre 9 et 11 études d'évaluation par année. La plupart des études sont toutefois concentrées sur des politiques, des programmes ou des initiatives individuels. En outre, lorsqu'ils sont soumis à l'approbation[1] du Comité ministériel de la vérification et de l'évaluation (CMVE), les rapports sont le plus souvent examinés individuellement, parce qu'ils sont présentés individuellement -- dès qu'ils sont terminés et à diverses périodes de l'année. Qui plus est, lorsque les rapports sont approuvés par le CMVE sans discussion formelle (par courrier électronique, par exemple), il arrive souvent que les problèmes qu'ils révèlent ne soient pas soulevés dans le cadre de considérations plus générales.

Il en résulte que les leçons qui se dégagent de l'évaluation d'une politique, d'un programme ou d'une initiative, les innovations signalées et les autres enjeux soulevés ne sont pas toujours largement partagés ou examinés aux fins d'une application plus générale en dehors du domaine évalué.

L'objet de la présente synthèse est de résumer les principales conclusions et recommandations tirées des études d'évaluation ayant été approuvées par le CMVE en 2005-2006 de façon à déterminer (et, s'il y a lieu, de généraliser) les leçons, innovations et autres informations présentées qui pourraient aider Santé Canada à améliorer l'efficacité de la gestion, des politiques ou des programmes dans d'autres domaines, y compris l'élaboration et la conception de nouvelles politiques et de nouveaux programmes. La synthèse permettra aussi d'examiner la conception des études d'évaluation et les processus qui y sont associés de manière à améliorer la contribution de cette fonction à la prise de décisions du Ministère.

Afin d'atteindre cet objectif, deux types de recherche – synthèse des évaluations et méta-évaluations – seront utilisées pour examiner de façon critique les rapports d'évaluation de 2005-2006.

---

[1] La politique de Santé Canada veut que toutes les évaluations effectuées au Ministère et les plans d'action de la gestion visant à répondre aux constatations, conclusions et recommandations découlant de ces évaluations soient soumis à l'examen et à l'approbation du Comité ministériel de vérification et d'évaluation (CMVE). Ce dernier surveille aussi la mise en oeuvre des plans d'action.

La synthèse d'évaluation est une façon systématique d'organiser les constatations tirées de plusieurs études d'évaluation.[2] L'évaluateur rassemble les résultats de divers rapports d'évaluation effectués par des personnes différentes à des périodes et à des endroits différents et « regroupe les constatations des études individuelles afin d'obtenir une conclusion plus crédible que celle de l'une ou l'autre des études ». [1, 3]

La **méta-évaluation** est une technique très semblable par laquelle on effectue un examen systématique des évaluations (et des évaluateurs) afin de déterminer la qualité des processus et des constatations.[4, 5] Une méta-évaluation sert à vérifier les évaluations par rapport à des problèmes tels que des parti pris, des erreurs techniques, des problèmes administratifs et juridiques et d'autres difficultés imprévisibles. Les méta-évaluations permettent aux organisations de s'interroger sur la clarté des constatations des évaluations, la pertinence du type d'évaluation choisi, le bien-fondé des conclusions par rapport aux données et la possibilité d'appliquer les résultats des études ou des recommandations qu'elles contiennent.

Le présent rapport se divise en quatre sections principales. L'introduction décrit brièvement l'objet de cet examen. La section sur les méthodes décrit la méthodologie employée. Elle est suivie d'une analyse des résultats de l'examen. La section finale résume les constatations et les conclusions de l'examen.

MÉTHODES

Onze évaluations ont été examinées par rapport à quatre attributs :

Fiabilité : La fiabilité représente la constance ou la stabilité d'une mesure, d'un test, d'un outil ou d'une technique. Elle est évaluée à partir de questions telles que : dans quelle mesure peut-on se fier aux résultats de l'évaluation et ces résultats sont-ils reproductibles ou simplement aléatoires, donc non significatifs? Des méthodes imparfaites de collecte de données et des erreurs d'échantillonnage peuvent résulter en une population non représentative aux fins d'une étude ou à des informations non fiables qui pourraient mener à des conclusions incorrectes.

---

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

3 General Accounting Office. Designing Evaluations. GAO/PEMD-10.1.4 Mars 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

Validité : Le travail porte ici sur la validité interne et externe. La *validité interne* est fonction de la véracité des constatations. Le type d'évaluation choisi permet-il de répondre à la question que celle-ci avait pour but d'élucider? Les résultats sont-ils exacts et non équivoques?

La *validité externe* se rapporte à la capacité de généraliser les conclusions de l'évaluation. À quelles populations et programmes peut-on appliquer les conclusions? L'information s'applique-t-elle seulement à l'échantillon d'où elle provient ou à d'autres groupes (donc, dans quelle mesure peut-on appliquer les conclusions à d'autres situations sans crainte de se tromper)?

Objectivité : L'objectivité se rapporte à la notion que les constatations et les conclusions sont fondées sur des faits obtenus par des méthodes ou un raisonnement scientifiques. Les résultats sont-ils hautement influencés par les expériences, les perceptions et les biais particuliers d'un évaluateur ou est-ce que, en général, d'autres évaluateurs, approuvent ou désapprouvent les conclusions? En règle générale, si des spécialistes différents, mais également compétents, ne peuvent arriver aux mêmes conclusions ou à des conclusions semblables au sujet des constatations d'une évaluation, les résultats de cette évaluation peuvent être en eux-mêmes remis en question.

Pertinence : Cet attribut est celui qui détermine si les constatations sont conformes à l'objet de l'évaluation. Les constatations de l'évaluation doivent correspondre à l'information que souhaitent obtenir ceux qui l'ont demandée pour que les résultats puissent être utilisés. La collecte de données qui peuvent servir à répondre correctement aux questions reliées à l'évaluation aide à résoudre les questions de pertinence.

ÉVALUATIONS APPROUVÉES PAR LE CMVE EN 2005-2006

Un nombre total de 11 évaluations ont été examinées aux fins du présent rapport. La plupart étaient des évaluations formatives, mais quelques-unes unes étaient des rapports sommatifs et des évaluations d'impact. Les évaluations et examens terminés et approuvés par le CMVE pour l'exercice financier 2005-2006 ont été les suivants :

1. Programme des partenariats pour l'infostructure canadienne de la santé (PPICS)
2. Initiative de recouvrement des coûts de l'Agence de réglementation de la lutte antiparasitaire (ARLA)
3. Programmes Grandir ensemble et Pour des communautés en santé de la Direction générale de la santé des Premières nations et des Inuits
4. Fonds d'innovation de Santé Canada
5. Programmes de subventions et de contributions dans le cadre des stratégies et des politiques en matière de soins de santé
6. Politique de transfert des services de santé aux Premières nations et aux Inuits
7. Fonds pour l'adaptation des services de santé

8. Protocoles conclus entre les sous-ministres adjoints et les directeurs généraux régionaux
9. Projets pilotes des Services de santé non assurés (SSNA) de la Direction générale de la santé des Premières nations et des Inuits
10. Fonds pour l'adaptation des soins de santé primaires
11. Division de la gestion et de la diffusion de la recherche

SYNTHÈSE DES CONSTATATIONS

Cette section de la synthèse fait ressortir les points forts et les faiblesses des programmes ou initiatives faisant l'objet de l'examen. Elle comprend les points suivants :

- les questions se rapportant à la mise en oeuvre;
- la question de savoir s'il y avait assez de faits probants pour tirer des conclusions sur l'impact ou les résultats pour la santé du programme ou de l'initiative en question, et
- des observations sur la qualité du travail d'évaluation lui-même.

PROBLÈMES DE MISE EN OEUVRE

Cinq problèmes importants ont été relevés par rapport à la mise en oeuvre des programmes.

1) Les objectifs, les buts et les critères opérationnels des programmes doivent être énoncés plus clairement

Dans les 11 rapports, il est signalé que, souvent, les objectifs et buts des programmes n'étaient pas assez clairs. Cela a mené à diverses interprétations sur la nature des résultats attendus, surtout dans les cas où de multiples paliers étaient concernés.

Quelques exemples :

- La rédaction des protocoles conclus entre les SMA et les DGR (rapport numéro 8) avait tendance à se situer à un niveau stratégique élevé et les protocoles ne comportaient pas de « procédures » ou « directives » formelles pour clarifier les rôles, les responsabilités, l'obligation de rendre compte et les particularités de leur mise en oeuvre à des niveaux plus opérationnels.

- Les critères d'administration de certains programmes, comme ceux qui sont le sujet du rapport numéro 3, ne sont peut-être pas assez clairs pour aider les gestionnaires des programmes communautaires à déterminer où et comment les fonds reçus devraient être affectés. Plus précisément, les critères relatifs aux types d'activités admissibles au financement des programmes Grandir ensemble et Pour des communautés en santé ne comportaient pas assez de directives pour que les gestionnaires des programmes des

Premières nations et des Inuits puissent, par exemple, savoir s'ils pouvaient utiliser les fonds de la composante du programme sur les abus de solvants pour financer d'autres types de programmes ou leur servir de complément dans le cas des collectivités où il n'y avait pas d'abus de solvants.

- Les objectifs et les critères opérationnels sont souvent déterminés aux premiers stades de l'élaboration d'un programme. Souvent des modifications peuvent être apportées au niveau de l'application du programme sans que des modifications correspondantes ne soient faites au niveau de l'autorité du programme, ce qui fait que les énoncés formels de politique et les directives ne reflètent plus ce qui se passe réellement. Le problème est encore aggravé lorsque de tels changements ne sont pas documentés (de la part des gestionnaires communautaires ou des gestionnaires régionaux des programmes de Santé Canada) et communiqués aux gestionnaires de Santé Canada au niveau des directions générales ou du programme national, et lorsqu'il y a lieu, uniformisés dans l'ensemble du programme.

- Souvent, les rôles et responsabilités ne sont pas clairement énoncés. Le rapport numéro 6, par exemple, conclut que l'incertitude relative aux rôles et responsabilités particulières de certaines collectivités des Premières nations et Inuits, des provinces et de la DGPNI en ce qui concerne les menaces à la santé publique pourrait faire que, dans l'éventualité d'une pandémie de grippe, les divers intervenants ne pourraient pas faire face efficacement ou de façon cohérente à la situation.

- Dans certains cas, on utilise une terminologie différente pour parler des mêmes concepts. Par exemple, dans le rapport numéro 5 on note que les objectifs des Programmes de subventions et de contributions dans le cadre des stratégies et des politiques en matière de soins de santé étaient parfois appelés « activités» ou « résultats » ou décrits comme tels. Des examens de la documentation des programmes ont aussi révélé que les objectifs énoncés n'étaient pas uniformes dans plusieurs des documents officiels sur la politique relative à ces programmes.

2) Retards relatifs au déboursement des fonds

L'affectation des fonds en temps opportun a été une question souvent soulevée dans certains des rapports. Cela était particulièrement le cas pour les initiatives et les programmes à délai d'application critique. Pour les programmes concernés par les rapports 4 et 10, par exemple, lorsque le financement a finalement été accordé pour certains projets, il ne restait pas assez de temps avant l'échéance des programmes pour les mettre en oeuvre effectivement et en évaluer l'impact. Le temps requis pour obtenir l'approbation ministérielle des accords de contribution était souvent la cause des retards.

3) Choix du moment où l'incidence des programmes est évaluée

Le temps requis pour évaluer l'efficacité des programmes est un élément relié à celui du déboursement des fonds. Les initiatives à délai prescrit comportent des exigences intrinsèques d'évaluation. Les retards de la mise en application ou du lancement des projets signifient que les évaluations doivent, si l'on veut qu'elles respectent les délais imposés par le Conseil du Trésor, avoir lieu trop tôt après le lancement du programme pour permettre de tirer des conclusions valables sur le rendement fondées des faits probants.

En outre, une période importante doit s'écouler avant que les résultats à long terme de certains programmes et projets puissent être observés -- une période parfois beaucoup plus longue que celle qui était couverte par leur financement. Tel était le cas du PPICS au sujet duquel le rapport numéro 1 concluait que l'évaluation, qui a été effectuée vers la fin du programme, n'a pas pu déterminer son impact à long terme.

En règle générale, dans le cas des programmes à délai prescrit, leur financement se termine souvent avant que les impacts à plus long terme ne puissent être démontrés ou observés. Par conséquent, le Ministère n'effectue pas le suivi au moyen d'évaluations pour déterminer si les résultats à long terme se sont en fait réalisés au moment prévu. Cette situation fait que lorsqu'on examine les réalisations découlant de ces initiatives à partir des faits probants disponibles, l'impact du travail de Santé Canada risque d'être qualifié de « non conclusif » ou « peu évident ».

4) Faiblesse de la mesure du rendement

Dans presque tous les rapports, on conclut que l'information sur le rendement est insuffisante. Dans certains cas, tels que celui du programme dont le rapport numéro 10 est l'objet, un cadre de mesure du rendement a été préparé, mais il n'a pas été mis en oeuvre. Dans d'autres cas, comme celui du programme évalué dans le rapport numéro 11, l'information disponible était d'une valeur limitée pour évaluer le rendement, car les données n'avaient pas été systématiquement recueillies ou catégorisées. Le rapport numéro 6 donne un exemple d'un programme où les collectivités concernées par l'initiative avaient recueilli l'information systématiquement, mais avaient utilisé des méthodes différentes de collecte et de regroupement de leurs données, les rendant disparates et difficilement comparables.

Quelques programmes, tels que celui dont il question dans le rapport numéro 5, avaient été dotés d'indicateurs et de mesures du rendement. Leur utilité était toutefois quelque peu compromise, parce que la plupart des indicateurs portaient sur les processus et les intrants et n'avaient que peu de liens directs avec les résultats attendus.

Dans le cas des initiatives communautaires concernant les Premières nations et les Inuits qui ont été évaluées en 2005-2006, la collecte de données et les rapports étaient orientés davantage sur les activités liées aux programmes et les transactions financières (c.-à-d. sur la conformité) que sur le rendement ou les résultats. Cela pourrait être attribuable au fait que Santé Canada a besoin

de retracer et de comptabiliser les fonds accordés pour les initiatives communautaires. Les gestionnaires de projets ont signalé que le nombre imposant de rapports qu'ils devaient soumettre au Ministère les obligeait à passer plus de temps et à consacrer plus d'énergie à la préparation de rapports mensuels sur les activités qu'à la collecte de données sur les résultats.

Une autre des complications entourant la mesure du rendement dans les programmes destinés aux Premières nations et aux Inuits est que, malgré que les composantes des programmes mis en oeuvre et des services dispensés par Santé Canada puissent être les mêmes, les communautés ont tendance à les administrer différemment. Il est compréhensible que l'on modifie les programmes pour répondre à des besoins uniques. Toutefois, dans la plupart des cas, lorsque ces modifications sont apportées, les ajustements correspondants aux résultats attendus ou aux buts du programme ne sont pas effectués. Même lorsque de tels ajustements sont faits, la tendance à s'attendre à ce que les programmes répondent aux objectifs originaux de la politique peut cependant dominer et mener ainsi à des constatations « non décisives ».

À un niveau plus général (et s'appliquant aux programmes concernant les Premières nations et les Inuits et à d'autres programmes), les programmes de subventions et de contributions sont souvent perçus par leurs bénéficiaires comme des sources de financement. Une fois les fonds reçus des diverses sources, ils sont souvent regroupés et affectés conformément aux « priorités locales », ce qui rend difficile, voire même impossible, l'établissement de liens entre le « rendement » et les « apports de fonds ».

5) L'échange d'information et le partage des connaissances entre les intervenants ne sont pas assez activement promus

L'une des conclusions du rapport d'évaluation numéro 1 indique que le programme aurait pu être plus réussi si les intervenants avaient communiqué entre eux les meilleures pratiques et les leçons qui se dégageaient de leur expérience. La question était particulièrement apparente dans le cas des programmes à délai prescrit comportant de multiples intervenants. Il a semblé que dans le cadre des objectifs d'administration des programmes, on n'a pas accordé d'importance à l'établissement de liens entre les projets de divers demandeurs et qu'on ne les ait pas encouragés à explorer la création de partenariats ou de réseaux afin d'améliorer les résultats découlant des programmes. Les pressions politiques voulant que les programmes deviennent « fonctionnels » aussitôt que les initiatives sont annoncées auraient, dans la plupart des cas, favorisé la considération des demandes individuelles aux dépens d'objectifs ou d'approches plus globales.

Les stratégies de diffusion relatives aux programmes de type recherche, tel que celui qui fait l'objet du rapport numéro 11, étaient généralement faibles. La plupart du temps, la diffusion des rapports et des conclusions de la recherche était fondée sur des stratégies passives plutôt qu'actives – p. ex. utiliser une stratégie générale plutôt que ciblée ou compter sur les chercheurs eux-mêmes pour faire la diffusion primaire.

Exemples de points forts de la mise en oeuvre

Bien que ce qui précède porte sur des déficiences clés de la mise en oeuvre des programmes, les évaluations et examens ont aussi révélé certains points forts.

Création de nouveaux outils : L'évaluation du PPICS a signalé le fait que ce programme avait réuni et présenté les résultats des projets et les évaluations comme éléments d'une base de données d'outils, de produits, de protocoles, de normes et de leçons à tirer. Cela a permis aux usagers d'accéder facilement à ces éléments d'apprentissage de façon à pouvoir les communiquer et à s'en servir comme point de départ d'autre recherche, s'ils le voulaient, afin d'accélérer l'avancement des technologies de l'information et des communications pancanadiennes.

Consultations inclusives : Les processus de consultation inclusifs, tels que ceux qui ont été utilisés pour le programme évalué dans le rapport 10, ont reconfirmé qu'en obtenant la participation de toutes les parties intéressées, on rehausse la manière dont les intervenants accueillent les programmes. Des efforts concertés ont été déployés pour obtenir la participation des groupes d'intervenants intéressés, y compris des provinces, des territoires et des organisations nationales non gouvernementales, dès le stade de la conception des programmes. Dans le cas du programme étant l'objet du rapport 10, il en est résulté la façon de mettre le programme en oeuvre était conforme à l'objectif de la réforme des soins de santé primaires voulant que les praticiens comprennent et fournissent les soins de santé dans les diverses régions du Canada selon une perspective holistique (plutôt que du seul point de vue du praticien).

PREUVES DE L'INCIDENCE DES PROGRAMMES

L'évaluation de l'incidence d'un programme consiste à déterminer dans quelle mesure ce programme a produit les changements souhaités dans une population cible. Elle comporte normalement une estimation ou une mesure des effets d'un programme, d'un projet ou d'une activité et par conséquent, exige souvent une analyse comparative de la situation avant et après leur mise en oeuvre.

En général, les 11 rapports d'évaluation et d'examen ont conclu que la capacité de Santé Canada de déterminer l'incidence de ses programmes est limitée par un manque de données sur le rendement, et dans certains cas, une absence de cadre de rendement. Les principales constatations à ce sujet ont été les suivantes :

1) Cadres analytiques requis pour l'évaluation d'impact

Les cadres analytiques (des critères d'évaluation de l'impact d'un programme sur lesquels on s'est entendu) étaient généralement absents. Un examen du rapport numéro 4, par exemple, révèle qu'il n'y avait pas de cadre permettant de déterminer comment l'impact du programme

serait mesuré. Sans un tel cadre, il est pratiquement impossible d'extraire tout effet prévu ou imprévu d'un programme. Dans de tels cas, la réussite ou l'échec est une conclusion subjective et discutable.

2) Absence de données de base

Les données de base[6] (le point de départ à partir duquel évaluer les résultats) sur les programmes n'existaient pas. La plupart des évaluations ont par conséquent dû reposer sur des analyses ex poste, qui ont mené surtout à des jugements qualitatifs subjectifs de la part du personnel chargé du programme ou des bénéficiaires. Les études évaluées se basaient largement sur les activités et les extrants pour déterminer l'incidence des programmes, plutôt que sur une approche plus rigoureuse et plus robuste, comme des comparaisons entre les données de base et celles qui suivent la mise en oeuvre.

3) Faiblesse des données sur les résultats

Les données sur le rendement des programmes étaient généralement absentes, limitées ou faibles dans toutes les études examinées. La principale raison en était que l'on n'accorde pas beaucoup d'importance à la collecte de données sur les résultats. La question est encore aggravée par le manque d'approches normalisées pour mesurer ces résultats. Par exemple, les projets évalués dans le cadre du rapport numéro 9 n'étaient pas comparables parce que les projets pilotes avaient été administrés différemment les uns des autres menant à des résultats qui ne pouvaient pas être généralisés aux autres communautés des Inuits et des Premières nations.

QUALITÉ DES RAPPORTS D'ÉVALUATION

L'évaluation d'un programme, d'une politique ou d'une initiative est une enquête systématique sur leur mérite ou leur valeur – et mène à des conclusions. Bien qu'elle soit souvent considérée, dans une certaine mesure, comme un « art », l'évaluation, en tant que discipline, comporte un ensemble de connaissances acquises et des méthodes généralement acceptées. La présente synthèse ne comprend pas d'examen systématique des processus et des méthodologies utilisées pour les 11 examens qui en font l'objet. Elle porte plutôt sur les rapports eux-mêmes, tels que rédigés, afin de déterminer les domaines possibles d'amélioration.

Quatre problèmes importants ont été relevés.

---

[6] Les données de base représentent l'information recueillie au départ sur un programme ou des éléments de programme avant que les services ne soient dispensés ou que l'on ait participé au programme. Elles sont souvent recueillies au moyen d'entrevues d'approche et d'observations et sont utilisées plus tard comme mesures de comparaison pour déterminer les changements apportés par un programme.

1) L'échantillonnage et la qualité des données pourraient être améliorés

La rigueur de la méthodologie est un élément essentiel de l'utilité des évaluations et de la possibilité de tirer des conclusions défendables. En outre, la capacité de généraliser du particulier au général améliore l'utilité de l'information. Toutefois, pour que les résultats d'une étude puissent être généralisables, il faut que l'échantillonnage soit représentatif du cadre d'échantillonnage (ou de la population cible de l'étude).

Les rapports d'évaluation examinés n'ont pas montré que les échantillons utilisés pour les études étaient représentatifs. Par conséquent, les évaluations ne peuvent permettre de conclure, à l'intérieur d'une marge d'erreur identifiable et déterminable, l'incidence ou la prévalence des constatations (tels que les résultats) sur les plans des politiques, des programmes ou des initiatives. En outre, la possibilité d'extrapoler les leçons qui se dégagent de telles évaluations à des initiatives semblables devient nulle.

Qui plus est, la qualité des données utilisées pour les rapports d'évaluation était généralement faible. Leur exactitude, leur cohérence, leur pertinence et leur intégralité étaient souvent discutables. Parallèlement, aucun des programmes évalués ne comprenait de normes de données vérifiables permettant d'assurer que celles qui étaient recueillies étaient exemptes de parti pris.

(2) Objectivité (ou « neutralité »)

Les études d'évaluation sont censées être des mesures du mérite, de la valeur ou des processus de gestion d'une politique, d'un programme ou d'une initiative fondées sur des faits probants (« indiscutables ») et « neutres ». La fiabilité et l'exactitude des constatations et des conclusions seraient améliorées si elles étaient confirmées par des données provenant de diverses approches ou « sources de données ».

Certaines des évaluations examinées ont utilisé de multiples sources de données, telles que la triangulation[7] dans leur recherche. En dépit de cela, il était évident que certaines sources « indépendantes » étaient affiliées de diverses manières au programme ou au domaine soumis à l'évaluation et que les conclusions étaient souvent principalement fondées sur des opinions subjectives. En outre, il était aussi possible que diverses sources de données (p. ex. groupes d'experts, analyses documentaires, entrevues avec des intervenants clés) se soient inspirées du même fonds d'information, tels que les intervenants spécialisés interviewés[8] rédigeant les

---

[7] La triangulation est une méthode d'établissement de l'exactitude de l'information qui consiste à comparer au moins trois types de points de vue indépendants sur les sources de données (par exemple, entrevues, observations et documentation de diverses périodes) et leur capacité d'en arriver à des conclusions identiques ou semblables.

[8] Tel qu'utilisé ici, le terme « intervenant » inclut Santé Canada.

documents examinés et agissant comme agents pour les organisations d'intervenants interviewées. L'objectivité des constatations fondées sur de telles entrevues, comme dans le cas du rapport numéro 6, peut donc être remise en question.

(3) Méthodes de recherche

Les rapports d'évaluation devraient inclure les décisions sur les méthodes ou approches principales utilisées qui permettraient au lecteur d'évaluer la validité et l'exactitude des constatations et conclusions. Par exemple, il pourrait être important de mentionner les critères d'inclusion et d'exclusion des sources de données.

Les critères d'inclusion et d'exclusion utilisés pour recruter les participants à des sondages ou à des entrevues étaient généralement absents des rapports. En outre, la justification des décisions d'inclure les gens dans les sondages ou les entrevues ou de les exclure était faible ou non existante. La description claire de ces paramètres pourrait permettre par exemple à des tiers qui le souhaitent de répéter facilement les échantillonnages et de tester la rigueur des constatations et des conclusions.

(4) Rapports

Les rapports d'évaluation doivent être cohérents; plus particulièrement, les recommandations doivent être fondées sur des constatations et conclusions découlant de faits probants et en étant la suite logique.

Dans certains rapports, sauf les rapports 4, 5, 8 et 11, les constatations ne correspondaient pas nécessairement aux conclusions. Les inférences de causalité décrites dans certains rapports n'étaient pas appuyées par des faits probants. On en trouve un exemple dans le rapport numéro 9 qui comprenait 41 recommandations non fondées sur des faits. Le *Guide d'évaluation des rapports d'évaluation* de la Direction de la mesure et de l'évaluation du rendement du Ministère (DMERM) a été utilisé comme outil pour en arriver à de telles conclusions.

EXAMEN DES PLANS D'ACTION DE GESTION

La politique de Santé Canada veut que chaque rapport d'évaluation comprenne un plan d'action de gestion. Ce plan décrit les mesures et processus que la direction hiérarchique s'est engagée à adopter pour aborder les enjeux présentés dans le rapport d'évaluation. (Il permet aussi aux gestionnaires hiérarchiques de présenter leurs perspectives, si elles sont différentes de celles de l'évaluateur.) Dans le cadre du processus d'approbation des rapports d'évaluation, la DMERM évalue le caractère raisonnable du plan d'action de gestion proposé par rapport aux

recommandations faites (et, le cas échéant, par rapport aux constatations et aux conclusions) et l'administrateur de la direction générale s'engage formellement à mettre le plan en oeuvre avant qu'il ne soit présenté à l'approbation du CMVE.

En règle générale, les plans d'action de gestion permettent d'aborder directement et adéquatement les enjeux soulevés dans l'évaluation. Toutefois, il arrive souvent que les intervenants doivent être consultés sur les détails des mesures particulières, surtout dans le cas des programmes de santé concernant les Premières nations et les Inuits. Ce fait est reconnu dans les plans présentés au CMVE, et la DMERM effectue le suivi auprès des domaines de programme concernés dans le cadre de ses responsabilités de surveillance de leur mise en oeuvre.

Deux problèmes surgissent par rapport aux plans d'action :

1) Les résultats des évaluations et les plans d'action de gestion subissent « un cloisonnement administratif »

L'un des avantages de l'évaluation se rapporte à l'application ou à l'adaptation des leçons qui s'en dégagent à des domaines qui n'y sont pas directement examinés. On a toutefois eu tendance à établir les plans d'action plus ou moins exclusivement en fonction des politiques, des programmes ou des initiatives évalués. Par conséquent, les mesures de gestion donnant suite aux constatations faites dans les évaluations, telles qu'une définition claire des objectifs, des buts et des critères opérationnels des programmes, ne semblent pas être considérées (ou perçues comme pertinentes) dans les domaines et les organisations qui n'étaient pas directement concernés par ces évaluations.

En effet, dans certains cas, les résultats de l'évaluation d'un programme à délai prescrit qui est remplacé par un programme légèrement différent (et peut-être portant un autre nom) ne semblent pas être considérés pour l'élaboration et la mise en oeuvre de ce dernier.

Par ailleurs, en dépit d'évaluations ayant fait ressortir régulièrement que l'information nécessaire pour appuyer des évaluations de qualité est souvent absente, surtout pendant les premières années des nouveaux programmes, de nombreuses présentations au Conseil du Trésor engagent encore le Ministère à mener de multiples évaluations au cours du cycle de vie typique de 5 ans des programmes à délai prescrit ou de subventions et contributions.

2) « Bonnes intentions » et programmes « éliminés »

En raison de l'orientation accrue sur les résultats et du fait reconnu qu'il faut un certain temps avant que les résultats de certains programmes de Santé Canada ne se manifestent suffisamment pour en faire une évaluation significative, certains gestionnaires de programmes éliminés (tels que le PPICS) ont engagé le Ministère à prendre des mesures après l'arrivée à terme du programme (et du financement).

Il s'agit là de bonnes intentions – p. ex, dans le cas du PPICS, l'intention consiste à évaluer l'impact des activités ministérielles au moment où il peut être mesuré de manière à conclure sur l'efficacité possible de la politique, et par là, à s'acquitter de l'obligation de rendre compte inhérente à la politique de Santé Canada – mais la probabilité que le suivi ait réellement lieu risque d'être mince, parce que l'organisation qui en serait responsable - ou le financement requis – pourrait ne plus exister au moment où le travail doit être fait.

CONCLUSION

Bien que la nécessité de faire des évaluations et d'en faire rapport au moment opportun soit reconnue, plus souvent qu'autrement, les prérequis d'une bonne évaluation, en particulier ceux qu'il faut pour évaluer l'efficacité et la rentabilité des activités, ont été absents. D'après les rapports d'évaluation examinés, s'il veut améliorer l'efficacité des programmes sur le plan des résultats et de leur gestion, ainsi que la qualité et l'utilité de la fonction d'évaluation elle-même, Santé Canada devrait considérer les points suivants :

- clarifier les objectifs, buts et critères opérationnels des programmes, et plus particulièrement, exiger l'établissement de cadres de gestion et d'analyse robustes qui seraient à la base d'une attribution claire des rôles et responsabilités et permettraient de déterminer et d'évaluer le rendement sur le plan de l'efficacité des programmes ou des résultats en matière de santé et de la gestion (processus);

- élaborer et mettre en oeuvre des initiatives de mesure du rendement – p. ex. cadres de rendement, orientation sur les résultats en sus des activités et des extrants, collecte et analyse des données – afin d'établir une riche base de données pour l'évaluation de l'efficacité des politiques, des programmes et de la gestion;

- améliorer la qualité des évaluations en en rehaussant la rigueur méthodologique et l'objectivité; et

- utiliser l'information sur les évaluations comme ressource « stratégique » – p. ex. évaluer les constatations, conclusions et recommandations des évaluations effectuées dans tout le Ministère – et non pas seulement pour la politique, le programme ou l'initiative qui en sont l'objet.