

*La conception des
évaluations sommatives
pour les prestations d'emploi et
mesures de soutien (PEMS)*

*(à l'intention du groupe d'experts
de Développement des
ressources humaines
Canada (DRHC) sur la conception des
évaluations sommatives pour
le programme des PEMS)*

*Rédigé par :
Walter Nicholson*

septembre 2001

Table des matières

Préface.....	i
A. Définir les « participants » aux fins des évaluations des prestations d'emploi et mesures de soutien (PEMS)	1
B. Méthodes d'analyse à utiliser dans les évaluations	5
C. Plan d'échantillonnage	17
D. Spécification et mesure des résultats obtenus par les participants	25
E. Coordination avec le projet des indicateurs à moyen terme (IMT)	29
F. Sommaire des questions sur lesquelles il faudra revenir au moment de la mise en œuvre.....	31
Annexe A : Évaluations sommatives des prestations d'emploi et mesures de soutien ainsi que des ententes sur le développement du marché du travail — répercussions éventuelles des programmes à examiner/mesurer	37

Préface

Développement des ressources humaines Canada (DRHC) et ses partenaires provinciaux et territoriaux reconnaissent l'importance des constats des évaluations lorsqu'il s'agit d'examiner les répercussions des prestations d'emploi et mesures de soutien ainsi que des ententes sur le développement du marché du travail (PEMS/EDMT). On trouve des dispositions concernant l'évaluation de ces éléments autant dans la législation sur l'assurance-emploi que dans les EDMT.

En général, l'évaluation des interventions en matière d'emploi constitue une tâche complexe. Dans le cas des évaluations sommatives des PEMS/EDMT, la grande diversité des collectivités et des modèles de mise en œuvre à l'échelle du pays, le nombre d'évaluations distinctes à réaliser et le calendrier d'exécution de chacune ajoutent à la complexité de la tâche.

Les évaluations appellent une certaine réflexion créative jumelée à une connaissance approfondie des techniques de mesure, notamment parce qu'elles doivent combiner la connaissance de théories compliquées du marché du travail et de l'économie, de modèles sophistiqués de mise en œuvre de programmes et d'acquisition du savoir, d'éventails complexes de répercussions éventuelles prévues et imprévues et de modèles statistiques fort structurés qui serviront à estimer, à partir de listes partielles de concepts, des résultats et des répercussions qui, bien souvent, ne peuvent être mesurés directement.

Par leur nature même, les évaluations des PEMS/EDMT nécessitent l'appel à de nombreux partenaires ayant divers intérêts, y compris les décideurs, les gestionnaires de programme, les agents de mise en œuvre, les consultants et les membres du public (comme les participants à un programme et les contribuables). Les membres de ces groupes possèdent des niveaux différents de connaissance des principes, des normes et des techniques d'évaluation ainsi que pour l'interprétation des constats. Pour que l'évaluation soit efficace, il est essentiel d'établir un niveau élevé de confiance entre tous les partenaires et toutes les parties intéressées. Il est donc important que la méthodologie utilisée soit fort justifiable et qu'elle soit présentée de façon claire et tout à fait transparente.

DRHC a confié à un groupe d'experts le mandat d'élaborer un cadre générique objectif et inspiré de l'état actuel des connaissances pour l'évaluation des PEMS/EDMT. Les experts ont élaboré le cadre de façon à offrir un fondement méthodologique commun et solide pour les diverses évaluations. Selon un des principes directeurs, il devait comporter assez de souplesse pour satisfaire aux exigences en matière de compte rendu aux niveaux national, provincial, territorial et local. Une de ses principales caractéristiques est qu'il est détaillé et qu'il peut être adapté facilement à des situations opérationnelles et à des contenus bien précis. La méthodologie proposée est fondée sur les plus récentes approches théoriques et pratiques pour l'évaluation ainsi que sur les points de vue et les expériences des experts en ce qui a trait à son application.

Le cadre proposé par le groupe d'experts est résumé dans le présent rapport. Cela devrait faciliter la conclusion d'une entente en ce qui concerne l'adoption d'une méthodologie, de sorte qu'au moment de communiquer les résultats, on pourra se concentrer sur l'interprétation des constats concernant les répercussions probables des PEMS/EDMT. Le rapport porte presque exclusivement sur les méthodes qui permettront de mesurer les résultats, plutôt que sur les éléments à mesurer. Le cadre d'évaluation proposé peut s'appliquer à un vaste éventail de résultats (notamment à ceux dont il est question à l'annexe A). DRHC tient à remercier M. Walter Nicholson, Jeff Smith, Chris O'Leary, Paul Decker et Joe Michalski de leur excellent travail.

DRHC présente ce rapport par souci de transparence en vue de faire progresser notre connaissance de la méthodologie des évaluations et faciliter les discussions à son sujet ainsi que pour assurer les meilleures évaluations sommatives possibles des PEMS.

Le présent rapport expose un cadre général qui peut constituer la base de toute évaluation bien conçue des prestations d'emploi et mesures de soutien (PEMS)¹. Il détermine les éléments qui devraient être considérés, passe en revue les points forts et les points faibles de diverses formules envisageables pour établir des mesures et souligne les questions qui pourraient faire l'objet de recherches supplémentaires².

Même s'il examine certaines questions de manière approfondie, ce document ne se veut pas le plan détaillé et définitif d'une évaluation en particulier. Comme on pourra le voir, le cadre général devra être adapté (modifié et élargi) afin de rendre compte de la singularité des exigences propres à chaque évaluation. Par exemple, tout plan d'évaluation définitif se fondera sur les buts de l'évaluation, la nature du programme à examiner, la qualité des données disponibles, de même que des considérations pratiques comme le calendrier de réalisation et les ressources disponibles.

La diffusion des résultats des évaluations et leur validité perçue représentent des éléments importants à prendre en considération dans l'évaluation des PEMS. Même si ce document tente de faire le point sur l'état actuel des connaissances relatives aux méthodes d'évaluation, il ne prétend pas définir ce qu'est une « bonne » évaluation, ni la façon dont les résultats obtenus par de telles méthodes devraient être utilisés. Pour être en mesure de juger de la validité globale des éléments de preuve, il faut particulièrement : 1) comprendre *a priori* l'applicabilité et les limites de l'approche générale utilisée (par exemple, les limites des modèles quasi expérimentaux); et 2) examiner *a posteriori* la qualité des données recueillies et la robustesse des méthodes d'estimation employées. C'est particulièrement important dans le contexte des évaluations des PEMS, car l'un des grands objectifs de la plupart de celles-ci sera probablement de recueillir des éléments de preuve pouvant aider à déterminer les « effets différentiels » des interventions. Même si des données bien comprises et bien structurées peuvent y contribuer grandement, de nombreux problèmes se posent lorsqu'on cherche à obtenir des estimations fiables des effets différentiels en contexte non expérimental. Dans beaucoup de cas, il pourra se révéler tout simplement impossible de produire des résultats valides à partir des données disponibles seulement. C'est pourquoi les évaluations planifiées doivent comprendre des procédures permettant d'analyser les points forts et les points faibles, de même qu'un processus clair pour l'intégration des résultats et leur diffusion auprès d'auditoires plus vastes.

Les mises en garde qui précèdent sont l'un des éléments qui justifient le cadre général présenté dans ce document, et elles y sont rappelées fréquemment. Elles sont nécessaires à une interprétation adéquate des suggestions du groupe d'experts. Pour interpréter et présenter adéquatement les éléments de preuve qui seront recueillis et qui pourront être diffusés à l'avenir, il faudra reconnaître les points forts et les points faibles possibles de diverses approches envisagées pour élaborer des évaluations sommatives.

¹ Les PEMS sont des « prestations d'emploi et mesures de soutien ». C'est la terminologie utilisée à l'échelle nationale, mais le nom du programme peut varier dans les régions.

² Ce rapport réunit, structure et développe une série de documents résumant les vues que le groupe d'experts sur les PEMS a exprimées lors d'une rencontre de deux jours (les 8 et 9 mars 2001 — voir Nicholson, 2001) et à l'occasion d'autres discussions générales ou plus pointues qui ont suivi.

Le document se divise en cinq sections analytiques qui examinent les éléments conceptuels généraux suivants :

- A. Définition des participants au programme des PEMS
- B. Méthodes d'analyse (y compris la sélection des groupes témoins)
- C. Plan d'échantillonnage
- D. Mesure des résultats obtenus par les participants
- E. Intégration au projet des indicateurs à moyen terme (IMT)

Une dernière section (F) résume certains des enjeux examinés et qui devraient être considérés au moment de la conception d'évaluations quasi expérimentales et comprend certaines lignes directrices sur la mise en œuvre des évaluations des PEMS.

A. Définir les « participants » aux fins des évaluations des prestations d'emploi et mesures de soutien (PEMS)

Pour concevoir une analyse claire des répercussions probables des interventions qu'englobe le programme des PEMS, il faut d'abord définir précisément la nature du programme et la façon dont ses « participants » sont déterminés. Deux facteurs généraux ont guidé l'examen de ces enjeux par le groupe d'experts : 1) la définition du programme et de ses interventions devrait être celle qui reflète le mieux la façon dont les services sont dispensés; et 2) la définition de la participation devrait refléter les expériences des individus, sans préjuger des résultats. Dans ce contexte, le groupe d'experts a fait les observations et les suggestions suivantes :

1. La définition des participants en fonction de la date de début du plan d'action semble être celle qui répond le mieux aux objectifs de l'évaluation. Le groupe d'experts est d'avis que c'est le concept du plan d'action qui reflète le mieux l'approche globale de la « gestion des cas » qui a été retenue dans le cadre du programme des PEMS. Même si les interventions dont profitent les participants sont assez différentes, la plupart des participants suivent le même processus d'inscription au programme, étant délibérément aiguillés vers les services jugés les plus avantageux. Par conséquent, le plan d'action (même s'il n'est pas utilisé explicitement dans toutes les régions) semble correspondre le plus étroitement à la nature du « programme » selon l'idée que s'en font les décideurs.

Il a paru au groupe d'experts que l'utilisation des dates de début des plans d'action pour identifier les participants au programme présentait divers avantages. Premièrement, il semble probable que les dates de début soient définies de façon relativement claire pour tous les participants. La solution de rechange qui consisterait à se concentrer sur les personnes qui ont cessé de participer au programme au cours d'une période donnée semble poser plus de problèmes, en partie parce que les dates de fin des plans d'action sont souvent définies de façon arbitraire. Une deuxième raison de préférer la date de début pour définir la participation est qu'il est probable qu'une telle définition corresponde mieux à d'autres données publiées, par exemple dans le cadre des Rapports de contrôle et d'évaluation. Finalement, la date de début semble représenter, d'un point de vue conceptuel, une ligne de démarcation plus nette entre les activités d'avant le programme et les « résultats obtenus par les participants », car, à certaines fins, tout ce qui se produit après la date de début peut être considéré comme un « résultat obtenu par les participants ». Ainsi, la période qui s'écoule pendant le programme est naturellement incluse et les coûts d'option afférents à la participation au programme peuvent être intégrés à l'évaluation générale.

Le groupe d'experts reconnaît évidemment que certaines régions n'ont pas recours à la formule de la gestion des cas/du plan d'action pour dispenser les interventions dans le

cadre des PEMS, et que la définition des participants en fonction des dates de début posera des problèmes conceptuels dans le cas d'interventions multiples ou lorsqu'il faudra faire, du point de vue des politiques, la distinction entre les résultats obtenus pendant et après le programme. Même si le groupe d'experts est convaincu que ces inconvénients possibles n'enlèvent rien à l'intérêt de la démarche qu'il propose (et qui semble suffisamment souple pour tenir compte d'un vaste éventail de procédures utilisées concrètement dans le cadre du programme³), il n'en croit pas moins que les concepteurs des évaluations à titre individuel devraient explorer ces questions à nouveau, en fonction des processus véritablement utilisés dans le cadre du programme et des données disponibles.

2. En général, les évaluations devraient se concentrer sur les personnes qui ont participé à un volet des « prestations d'emploi ». Les quatre prestations d'emploi (« PE » — c'est-à-dire Subventions salariales ciblées, Aide au travail indépendant, Partenariats de création d'emplois et Développement des compétences) sont au cœur des PEMS. Elles sont également les interventions les plus coûteuses par participant. Le groupe d'experts est donc d'avis que les évaluations devraient se concentrer sur les participants à ces interventions. Les régions voudront probablement évaluer les répercussions de chacune de ces interventions séparément. Cela représentera des problèmes particuliers au chapitre de la conception des évaluations. Nous y reviendrons dans la section C ci-dessous, où nous montrerons que les modèles optimaux permettant de répondre à cet objectif qui consiste à évaluer des interventions à titre individuel pourront différer suivant les intérêts particuliers des décideurs régionaux. De plus, selon le groupe d'experts, il y aurait lieu d'envisager d'inclure un groupe distinct de clients participant uniquement aux mesures de soutien dans certaines évaluations. Cette possibilité est examinée ci-dessous.

3. La « participation » appelle une définition précise — peut-être selon le financement (si c'est possible). Le groupe d'experts se préoccupe à l'idée qu'une date de début d'une PE en particulier (ou d'un plan d'action en général) puisse avoir été attribuée à certains clients des PEMS sans qu'ils n'y aient vraiment pris part. En supposant que de tels cas sont relativement communs (il faudrait cependant recueillir des données à ce sujet), le groupe d'experts est d'avis qu'il devrait y avoir participation véritable, même minimale, à une intervention. Une possibilité serait de définir la participation en fonction du financement observé qui a été consacré à un client ayant participé à l'intervention en question. Il faudra cependant procéder à d'autres recherches pour déterminer s'il existe des données sur le financement individuel pour chacune des interventions PE. Il semble probable que les définitions de la participation reposant sur le financement différeront d'une région à l'autre, parce que les façons dont ces données sont recueillies différeront elles aussi.

³ Par exemple, dans les régions qui n'ont pas retenu explicitement la formule de la gestion des cas/des plans d'action, les dates de début et de fin des interventions pourraient tenir lieu de dates de début et de fin des plans d'action. Pour déterminer précisément la façon de procéder à cette simulation, il faudra connaître à fond les politiques et les pratiques de mise en œuvre au niveau local. Il est possible que les définitions doivent varier d'une région à l'autre pour que les plans simulés reflètent fidèlement la façon dont le programme est véritablement dispensé. C'est une question que les comités conjoints d'évaluation et Développement des ressources humaines Canada (DRHC) devront examiner soigneusement lorsqu'ils concevront la version finale de leurs évaluations provinciales/territoriales et qu'ils feront connaître les résultats à l'échelle nationale.

4. L'achèvement du programme n'est pas un bon critère de sélection des membres de l'échantillon. Bien qu'il puisse paraître juste de faire porter l'évaluation des PEMS sur les seuls participants qui ont achevé le programme, le groupe d'experts pense que cette approche ne serait pas de mise. L'achèvement d'une PE est en soi un résultat présentant de l'intérêt. Il ne faudrait pas en faire une condition pour la sélection de l'échantillon.

5. Dans certaines régions, il conviendrait d'envisager une cellule distincte pour les clients des services d'aide à l'emploi (SAE) seulement. Le programme des PEMS consacre environ le tiers de son budget aux mesures de soutien, et celles-ci devraient être évaluées séparément. Des recherches antérieures ont montré que des interventions relativement modestes sur le plan de l'emploi peuvent afficher le meilleur rapport coût-efficacité (voir la section C). Le groupe d'experts croit donc que si l'on exclut de l'analyse tous les individus participant uniquement à une intervention MS, on risque de passer à côté d'une bonne partie de ce qui fait l'intérêt du programme global des PEMS. C'est particulièrement vrai dans les régions qui mettent l'accent sur les mesures de soutien. En outre, il semble possible qu'une cellule composée uniquement de participants aux SAE puisse contribuer à l'estimation de l'effet des interventions PE en soi, parce qu'un échantillon exclusivement SAE peut, dans certaines circonstances, se révéler un bon groupe témoin pour certaines interventions PE. De nombreuses autres évaluations ont adopté une approche à plusieurs niveaux pour définir les traitements, selon laquelle les traitements plus complexes représentent des « ajouts » aux traitements plus simples. Il se peut en fait que, dans certains endroits, les PEMS soient exploitées selon cette approche à plusieurs niveaux. Les rôles éventuels que pourrait jouer une cellule composée uniquement de participants aux SAE sont examinés dans les sections B et C; le groupe d'experts pense que la plupart des évaluations devraient envisager cette possibilité.

6. Les apprentis devraient faire l'objet d'une étude distincte. Même si la prestation de services aux apprentis est une composante importante du programme des PEMS, le groupe d'experts est d'avis que les questions méthodologiques qu'il faut régler avant de pouvoir étudier ce groupe adéquatement nécessiteraient un plan de recherche distinct. Parmi les grandes questions que soulèverait nécessairement une telle étude, on retrouverait probablement celles-ci : 1) Comment définir les périodes de participation au programme des apprentis? 2) Comment établir un groupe témoin pour les apprentis — est-il possible de trouver des individus présentant des « liens avec le marché du travail » similaires? 3) Comment les résultats obtenus par les apprentis devraient-ils être définis? Comme les réponses qui pourraient être apportées à ces questions ne correspondent pas exactement aux thèmes examinés dans les ouvrages plus généraux sur l'emploi et la formation, le groupe d'experts pense que le simple fait d'ajouter un « traitement » des apprentis dans le plan général d'évaluation sommative apporterait peu d'informations utiles. Cela détournerait en outre certaines ressources précieuses de la réalisation d'autres objectifs d'évaluation.

B. Méthodes d'analyse à utiliser dans les évaluations

Le groupe d'experts reconnaît que l'évaluation des effets différentiels des PEMS constitue un objectif opportun⁴, mais que si on cherche à évaluer ces effets dans un contexte non expérimental (c'est-à-dire en faisant appel à des modèles quasi expérimentaux), il faudra relever des défis importants. La mesure de tels « effets différentiels » sous forme d'estimations définitives des répercussions d'une intervention sur les participants par rapport aux groupes témoins soulève d'importantes questions méthodologiques. Dans beaucoup de cas, compte tenu des données disponibles, il peut tout simplement être impossible de trouver des éléments de preuve définitifs au sujet des « vraies » répercussions. Les résultats des évaluations devraient alors, dans certaines conditions clairement définies, être présentés comme des « indicateurs de répercussions possibles ». L'extension des résultats à des situations qui ne correspondent pas à ces conditions devrait être soumise à un examen attentif et critique.

Les enjeux peuvent être illustrés comme suit : un objectif des évaluations sommatives est de produire des estimations « cohérentes » et relativement « précises » des résultats chez les participants et les membres du groupe témoin. En d'autres termes, la méthode devrait donner des estimations qui, si les échantillons étaient très vastes, se rapprocheraient des valeurs vraies (pour l'ensemble de la population). Les estimations devraient aussi afficher une faible variabilité d'échantillonnage, de sorte qu'il soit peu probable qu'elles s'écartent beaucoup des valeurs vraies. En fait, les gestionnaires s'attendent souvent à ce que les évaluations puissent produire des estimations cohérentes et précises de « l'effet différentiel » (c'est-à-dire l'effet du programme sur les résultats par rapport à ce qui se serait produit en l'absence de l'intervention). Le groupe d'experts reconnaît qu'il sera difficile d'atteindre ces buts au moyen d'un modèle quasi expérimental. En effet, un tel modèle comporte inévitablement le problème fondamentalement insoluble de la comparabilité des participants et des témoins, en particulier lorsque les différences entre les groupes sont liées à des caractéristiques qui ne sont pas directement observées lors de l'évaluation. Le groupe pense qu'il n'est pas possible de préciser *a priori* la « meilleure » formule qui se révélera optimale pour produire des estimations adéquates dans toutes les circonstances. Il recommande donc que les évaluateurs adoptent une démarche vaste et variée, explorant diverses méthodes, tant pendant la phase de planification que pendant la phase d'analyse de leurs travaux. De plus, le groupe d'experts est fermement convaincu que les approches retenues devraient être soigneusement étayées et faire l'objet de comparaisons critiques. Il croit également qu'il faudrait encourager les évaluateurs à déterminer la meilleure démarche à adopter.

⁴ D'autres objectifs pourraient se révéler importants pour certains décideurs, et notamment : 1) obtenir de bonnes données sur les caractéristiques des participants; 2) étudier le processus de mise en œuvre des interventions des PEMS; et 3) mesurer les expériences « brutes » des participants après le programme. Même si la plus grande partie de ce rapport (et des délibérations du groupe d'experts) a porté sur les moyens d'obtenir des estimations des effets différentiels, ces autres objectifs peuvent se révéler assez importants dans certaines régions et ils joueront donc un rôle de premier plan dans la conception des évaluations pour ces régions.

1. Stratégies envisageables pour la mesure des résultats

Selon le groupe d'experts, diverses approches semblent envisageables pour mesurer les résultats obtenus par les participants et les non-participants aux interventions des PEMS dans le cadre des évaluations sommatives. Pour que rien ne soit négligé, le groupe a dressé une liste assez complète de possibilités, qui sont énumérées ici, chacune étant commentée.

- a. Les méthodes de répartition aléatoire (« méthodes expérimentales ») :** La répartition aléatoire demeure la « règle d'or » dans les évaluations relatives au marché du travail. Elle garantit une estimation cohérente et efficace de l'effet moyen d'une intervention sur le participant⁵ et est devenue la norme de comparaison pour toutes les autres approches (voir, par exemple, Lalonde, 1986, ainsi que Smith et Todd, 2001). Compte tenu de ces avantages, le groupe d'experts est fermement convaincu qu'il conviendrait d'envisager la possibilité de recourir à une forme ou une autre de répartition aléatoire dans les évaluations sommatives. Le groupe reconnaît évidemment que le principal inconvénient de la répartition aléatoire est qu'elle n'est pas compatible avec l'objectif fondamental de l'accès universel du programme des PEMS. Dans la mesure où les personnes faisant partie d'un groupe témoin ne pourraient pas participer au programme (du moins pendant un certain temps), cela placerait les besoins de l'évaluation en contradiction avec les principes sur lesquels repose le programme. La façon habituelle de régler de tels conflits est de concevoir des interventions améliorées par rapport aux interventions normalement offertes, selon le principe voulant qu'il soit moralement moins répréhensible de refuser à un individu de participer à l'intervention améliorée. Si le financement prévu pour certaines interventions est limité, ces interventions elles-mêmes pourraient être considérées comme des « améliorations », de sorte qu'en pareil cas, la répartition aléatoire serait également justifiable. D'autres contraintes (en ce qui concerne la capacité du programme, par exemple) pourraient également servir de critères à la répartition aléatoire dans les évaluations, de manière à ne pas entraver l'admissibilité universelle aux programmes.

Toutefois, compte tenu des constats tirés de la mise en œuvre du programme des PEMS sur le terrain, le groupe d'experts conclut de façon générale que de telles options ne sont pas fréquentes dans le cadre du programme tel qu'il est actuellement conçu. Le groupe est néanmoins d'avis que les évaluateurs devraient toujours envisager en premier lieu la possibilité de mener une évaluation reposant sur une répartition au hasard, avant de se tourner vers des solutions de second choix. Les évaluateurs devraient également indiquer les cas où les évaluations reposant sur une répartition aléatoire pourraient se révéler fort utiles lorsqu'il s'agit de clarifier des constats ambigus devant servir à la détermination des répercussions. On pourrait ainsi cerner les cas dans lesquels la répartition aléatoire pourrait être utilisée fort efficacement pour compléter d'autres recherches visant le programme des PEMS.

⁵ Il conviendrait de modifier cet énoncé si le traitement expérimental visait également les membres du groupe témoin (par exemple, si les participants étaient placés au premier rang dans les files d'attente). On trouvera un examen de certaines questions plus complexes concernant les conclusions qu'on peut légitimement tirer des modèles de répartition aléatoire dans Heckman, Lalonde et Smith, 1999.

Il faudrait aussi considérer la répartition au hasard à des fins autres que la création de concepts de traitement-contrôle additionnels. Une gamme d'améliorations d'importance croissante aux programmes pourraient être assignées au hasard et s'ajouter à une option de service universelle. En outre, la répartition au hasard pourrait permettre de créer des « outils » utiles pour identifier la participation au programme dans le cadre d'évaluations qui s'avèreraient autrement non expérimentales. Pour réaliser ceci, on pourrait imposer des exigences de counselling additionnelles à un groupe de prestataires de l'AE sélectionnés au hasard par exemple, ce qui pourrait entraîner une variation fortuite au niveau de la formation. On pourrait réaliser un autre objectif similaire en misant sur des primes monétaires allouées au hasard. L'essentiel, c'est qu'une répartition au hasard innovatrice pourrait s'avérer utile de maintes façons en vue d'améliorer la qualité des évaluations de PEMS. Compte tenu de la validité additionnelle qu'on n'obtient souvent qu'en utilisant la répartition au hasard, le groupe d'experts appuierait avec enthousiasme des efforts si novateurs.

b. Les méthodes non expérimentales : De nombreuses méthodes non expérimentales ont été utilisées pour l'évaluation des programmes du marché du travail lorsqu'une répartition aléatoire est impossible. La plupart ont recours à des groupes témoins⁶, dans l'espoir que les expériences des témoins représenteront fidèlement la situation dans laquelle se seraient retrouvés les participants s'ils n'avaient pas participé au programme. Nous présenterons d'abord une liste relativement simple des possibilités. Nous verrons ensuite de façon plus détaillée, dans la sous-section suivante, de quelle façon les méthodes reposant sur des groupes témoins interagissent avec les méthodes de mesure lorsqu'il s'agit de déterminer s'il est possible de produire des estimations cohérentes des résultats obtenus par les participants. Par ailleurs, comme nous le démontrons, l'efficacité des méthodes pourra varier, selon que ce sont des facteurs observables ou non observables qui expliquent les différences entre les participants et les témoins; nous utilisons donc cette distinction pour illustrer les différentes approches.

i. Méthodes qui tiennent compte des différences observables. De telles méthodes sont faciles à appliquer, mais puisqu'elles ne tiennent pas compte des différences inobservables dans les déterminants de la participation au programme, la cohérence des estimations qu'elles produisent⁷ demeure douteuse. Cependant, selon certains auteurs, les stratégies d'appariement peuvent se révéler très efficaces pour constituer des groupes témoins valables (Rosenbaum et Rubin, 1983, Dehejia et Wahba, 1998, 1999), de sorte que le groupe d'experts préconise qu'elles soient envisagées parmi les méthodes de mesure des résultats dans le cadre des évaluations des PEMS :

⁶ Le recours à la méthodologie avant-après n'exige pas de groupe de comparaison explicite. Les extrapolations tirées des comportements passés jouent ce rôle.

⁷ Il peut y avoir incohérence, par exemple, si une variable non mesurée (notamment la « motivation ») influence à la fois la participation au programme et les résultats sur le marché du travail. Si les personnes plus motivées sont plus susceptibles de participer et qu'elles obtiennent également des résultats plus favorables sur le marché du travail lorsqu'il n'y a pas de participation, la participation au programme semblera « expliquer » ces résultats favorables.

- **Comparaison des moyennes rajustées par les moindres carrés ordinaires (MCO).** Ces estimations sont généralement utilisées pour faire une analyse préliminaire des données. La méthode donne des estimations relativement valables qui tiennent compte (de façon linéaire) des différences observables entre les participants et les membres du groupe témoin. Ces résultats peuvent contribuer à clarifier la nature des données, mais ne sauraient être considérés comme des estimations définitives des répercussions, surtout parce que la méthode n'offre aucune garantie contre l'influence que peuvent avoir des différences non mesurées, comme celles qui relèvent de facteurs inobservables.
- **MCO avec variables décalées des résultats.** Ces estimations présentent des problèmes semblables à ceux des estimations simples par les MCO. Dans certains cas, on peut améliorer la situation en tenant compte des résultats décalés, ce qui permet de prendre en considération une partie des différences inobservées entre participants et membres du groupe témoin⁸. Cependant, on risque ainsi d'introduire des biais dont on ne connaît pas le sens et les estimations qui en résultent peuvent être très sensibles à la façon dont les variables décalées ont été spécifiées.
- **Méthodes reposant sur l'appariement.** Il existe diverses stratégies d'appariement qui pourraient être mises en œuvre dans les évaluations sommatives des PEMS⁹. Comme nous le verrons dans la section qui suit, elles pourraient reposer uniquement sur des données administratives, ou encore sur une combinaison de données administratives et de données d'enquête. Même si l'appariement ne tient pas directement compte des différences non mesurées entre participants et membres du groupe témoin, il peut le faire de façon approximative si les facteurs inobservables sont corrélés aux variables utilisées pour l'appariement¹⁰. Comme nous le verrons dans la section C, l'adoption de procédures d'appariement aurait également des conséquences pour la répartition de l'échantillon — notamment parce qu'elle aurait pour effet d'accroître la taille des groupes témoins afin de tenir compte de la possibilité que certains des membres de l'échantillon initial ne soient bien appariés à aucun participant. L'adoption de stratégies d'appariement poserait également des problèmes de logistique au moment de la collecte des données d'enquête aux fins de l'évaluation, et on y reviendra dans la section C.

⁸ Par exemple, comme la « motivation » influence les résultats antérieurs aussi bien que les résultats futurs sur le marché du travail, on peut, en tenant compte des résultats antérieurs, obtenir une mesure (imparfaite) de la motivation.

⁹ En général, les méthodes d'appariement peuvent être jugées supérieures aux procédures MCO lorsqu'il s'agit de contrôler les variables observables parce que cette méthode n'entraîne pas d'hypothèse linéaire implicite aux MCO.

¹⁰ Techniquement, la qualité des données utilisées au niveau de l'appariement devrait faire en sorte que la participation au programme soit indépendante des résultats non traités, conditionnels aux variables utilisées pour procéder à l'appariement.

On retrouve deux formules générales d'appariement dans la littérature :

- **L'appariement aux caractéristiques:** Ces procédures utilisent soit un appariement exact entre des cellules fondé sur les caractéristiques du participant ou une sorte d'algorithme de distance qui apparie les membres du groupe témoin¹¹ à des participants donnés, selon des caractéristiques observables. Les estimations peuvent différer beaucoup selon les caractéristiques utilisées au niveau du concordeur. Dans certains cas, les chercheurs ont tenté de simuler les incertitudes que suppose l'appariement exact en ayant recours à plusieurs applications différentes des algorithmes d'appariement.
- **L'appariement probabiliste :** Ce procédé (dont Rosenbaum et Rubin, 1983, ont été les pionniers) effectue l'appariement à l'aide de la « propension » prédite à participer à un programme. Il faut d'abord faire une estimation de premier degré des probabilités de participation en fonction de caractéristiques observables. La procédure peut donner de bons résultats si deux hypothèses additionnelles¹² se concrétisent : 1) la participation peut être « très précisément » prédite ; et 2) la distribution des probabilités prédites est « suffisamment semblable » entre le groupe des participants et le groupe témoin. Dans la pratique, l'appariement probabiliste peut aussi poser des difficultés de mise en œuvre, parce que les déterminants de la participation ne sont pas bien connus et que des spécifications différentes de l'équation de la participation peuvent donner des appariements assez différents. (Les chercheurs devraient énumérer, expliquer et étayer soigneusement les critères d'utilisation des expressions « très précisément » et « suffisamment semblable »).

ii. Méthodes qui tiennent compte des différences inobservables

Ces méthodes reposent également sur des hypothèses au sujet de la nature des différences inobservables entre les participants et les non-participants. S'appuyant sur ces postulats, elles tentent d'expliquer et d'éliminer ces différences *a priori* par la soustraction, la modélisation complexe ou une sélection minutieuse des groupes témoins à partir de caractéristiques détaillées. La validité des estimations dépend de la qualité des données disponibles, du bien-fondé des postulats concernant les facteurs inobservables ainsi que de la robustesse des techniques. Il

¹¹ Selon les deux formules d'appariement, les cas de comparaison sont choisis « avec remplacement » — c'est-à-dire qu'un cas de comparaison peut parfois être le plus étroitement apparié à deux participants ou plus. On pense généralement que la réduction d'efficacité que représente cette double utilisation est plus que compensée par un meilleur appariement. Notons cependant que cette méthode ne serait pas adéquate si le bassin des non-participants était si limité par rapport à celui des participants que de nombreux cas seraient tirés à répétition (par exemple, la taille du bassin des candidats à l'appariement ne devrait pas être de plus de deux ou trois fois inférieure à celle du groupe de participants). Généralement, les procédures d'appariement ne fonctionneront pas bien si l'ensemble des caractéristiques d'appui font état d'une grosse différence entre la comparaison et les groupes participants.

¹² C'est à dire en plus de l'hypothèse générale nécessaire à toutes les procédures d'appariement à l'effet que la participation au programme doit être indépendante des résultats non traités, compte tenu des variables observables.

est donc important que les évaluateurs, au moment d'évaluer les répercussions des PEMS, prennent en compte les dérogations aux postulats sous-jacents et leurs effets possibles sur les estimations correspondantes.

Les méthodes qui tentent de tenir compte des différences inobservables incluent :

- **Les méthodes des écarts dans les différences.** Ces méthodes reposent sur l'hypothèse cruciale voulant que les différences inobservables entre les individus soient constantes dans le temps pour chaque individu — de sorte qu'elles disparaissent dans les données qui concernent chacun au moment où sont calculées les différences attribuables aux facteurs observables. Avec un nombre suffisant d'observations antérieures ou postérieures au programme, il est possible, en principe, de tester cette hypothèse en ayant recours à des périodes de temps autres pour mesurer les différences antérieures et postérieures. Cependant, si les différences inobservables changent au fil du temps ou sont influencées par la participation au programme, la méthode des écarts dans les différences ne donnera pas d'estimations cohérentes.
- **Méthodes Heckman/VI.** Ces méthodes reposent sur l'existence d'une « variable instrumentale » (VI) devant satisfaire deux critères : 1) être indépendante du résultat mesuré; et 2) avoir une valeur prédictive significative pour expliquer les décisions relatives à la participation au programme. Les « bons » instruments sont plutôt rares, et les évaluateurs qui adoptent ces techniques devraient se montrer prudents. Dans certains cas, on peut générer des instruments lors d'une évaluation¹³. Le groupe d'experts est d'avis que, lorsqu'ils présentent les résultats obtenus selon ces méthodes, les évaluateurs devraient indiquer clairement l'instrument qui a été utilisé et fournir des tests relatifs aux spécifications qui permettront de déterminer si la variable répond aux critères nécessaires. Les évaluateurs devraient également envisager la possibilité de générer des variables instrumentales, lorsque c'est possible.

2. Groupes témoins, méthodes d'estimation et cohérence

Les techniques décrites dans la sous-section précédente ont souvent été utilisées de diverses façons dans les évaluations. Dans certains cas, une procédure est adoptée *a priori* et aucune autre n'est utilisée. Dans d'autres cas, les chercheurs adoptent une démarche plus éclectique, dans le cadre de laquelle ils recourent à une méthode d'appariement, font ensuite une analyse de régression puis, pour faire bonne mesure, appliquent certaines techniques VI. En général, le groupe d'experts estime que ni l'une ni l'autre de ces démarches n'est idéale. La sélection d'une seule démarche analytique dans le cadre d'une évaluation pourra souvent se révéler inadéquate une fois les données examinées en détail. D'autre part, une démarche hétéroclite peut donner lieu à l'adoption de techniques

¹³ Par exemple, l'information découlant des évaluations faites par le personnel au sujet des candidats lorsqu'il s'agit de déterminer s'ils sont aptes ou non à participer à un programme peut servir de variable instrumentale de ce genre.

incompatibles les unes avec les autres. Ce qu'il faut plutôt, c'est une justification détaillée de toutes les approches, quelles qu'elles soient, qui ont été retenues pour une évaluation en particulier, parallèlement à une évaluation *a posteriori* de la validité vraisemblable des approches retenues.

Afin de clarifier les questions ayant trait à la sélection des approches analytiques aux fins des évaluations, le tableau 1 examine de façon assez détaillée l'interaction entre le choix du groupe témoin et la méthode d'estimation. On y envisage trois possibilités quant au choix du groupe témoin, selon la façon dont les membres seront appariés¹⁴ aux participants :

1. appariement selon une série limitée de variables administratives (par exemple, les données de l'assurance-emploi (AE) seulement), dites V1. Pratiquement toutes les évaluations, peu importe leur modèle, auront recours à cette méthode d'appariement — ne serait-ce que pour comparer les dates auxquelles commencent les périodes de versement des prestations aux prestataires;
2. appariement selon des variables V1 et d'autres variables administratives (V2 — qui, en principe, comprendraient la rémunération antérieure et l'impôt sur le revenu selon l'ADRC, car elle peut se révéler l'information administrative la plus instructive). Cela supposerait un appariement assez exhaustif à partir des données administratives avant que les enquêtes soient entreprises;
3. appariement selon des variables V1, V2 et d'autres variables que l'on peut recueillir seulement dans le cadre d'enquêtes, V3 (comme des données sur la rémunération familiale récente ou le processus à l'issue duquel les répondants se sont inscrits/ne se sont pas inscrits au programme des PEMS).

Chacune de ces méthodes de sélection des groupes témoins est corrélée aux quatre approches qui pourraient être envisagées pour générer les estimations des différences dans les résultats obtenus par les participants et par les non-participants :

1. écarts dans les moyennes;
2. écarts dans les moyennes avec rajustement par les MCO;
3. écarts dans les différences (avec rajustement par les MCO);
4. rajustement par variables instrumentales (VI) (y compris les « méthodes de Heckman »).

¹⁴ Il ne s'agira pas nécessairement d'appariements par paires. On pourrait aussi inclure des méthodes où plus d'un non participant pourrait servir à bâtir une « comparaison » pour chaque participant (voir Heckman, Ichimura et Todd, 1997).

TABLEAU 1				
Cohérence des estimateurs dans diverses circonstances				
Appariement du groupe témoin selon	Cohérence¹⁵ si les participants et les membres du groupe témoin diffèrent selon			
	V1 seulement	V1 ou V2	V1, V2 ou V3	V1, V2, V3 ou U¹⁶
Écarts simples dans les moyennes				
V1	Oui	Non	Non	Non
V1, V2	Oui	Oui	Non	Non
V1-V3	Oui	Oui	Oui	Non
Écarts dans les moyennes avec rajustement par régression***				
V1	Oui	Oui*	Oui*	Non
V1, V2	Oui	Oui	Oui*	Non
V1-V3	Oui	Oui	Oui	Non
Estimations des écarts dans les différences				
V1	Oui	Oui*	Oui*	Oui — si U ne varie pas dans le temps. Non autrement.
V1, V2	Oui	Oui	Oui*	Oui — si U ne varie pas dans le temps. Non autrement.
V1-V3	Oui	Oui	Oui	Oui — si U ne varie pas dans le temps. Non autrement.
<p>* Le rajustement par les MCO produira ici des estimations cohérentes, mais moins efficaces que celles qu'on obtiendrait si l'appariement avait reposé sur toutes les variables mesurables. Il importe de noter que l'appariement reposant sur toutes les variables disponibles produirait généralement des estimations cohérentes, alors que le rajustement par les MCO pourrait ne pas accomplir ceci en raison de l'hypothèse de linéarité inhérente aux MCO.</p> <p>** L'emploi d'algorithmes exhaustifs d'appariement peut influencer l'efficacité des méthodes d'estimation VI.</p> <p>*** On tient pour acquis que les régressions de MCO comprendraient les trois types de variables mesurables, alors que l'appariement ne serait fondé que sur les ensembles de variables identifiées dans le tableau.</p>				

¹⁵ On fait référence à la cohérence statistique comme on l'a décrite plus tôt (si les échantillons étaient très vastes, les estimations se rapprocheraient des valeurs vraies pour l'ensemble de la population).

¹⁶ U désigne des variables inobservables — il s'agit de variables qui peuvent influencer les résultats tout comme la participation au programme, mais qui ne sont pas observées, ni dans les données administratives, ni dans les données d'enquête (par exemple la motivation).

TABLEAU 1 (suite)				
Cohérence des estimateurs dans diverses circonstances				
Appariement du groupe témoin selon	Cohérence si les participants et les membres du groupe témoin diffèrent selon			
	V1 seulement	V1 ou V2	V1, V2 ou V3	V1, V2, V3 ou U
Estimateurs VI				
V1	Oui**	Oui**	Oui**	Oui si les VI sont bonnes et que la méthode n'est pas compromise par l'appariement ¹⁷
V1, V2	Oui**	Oui**	Oui**	Oui si les VI sont bonnes et que la méthode n'est pas compromise par l'appariement
V1-V3	Oui**	Oui**	Oui**	Oui si les VI sont bonnes et que la méthode n'est pas compromise par l'appariement
<p>* Le rajustement par les MCO produira ici des estimations cohérentes, mais moins efficaces que celles qu'on obtiendrait si l'appariement avait reposé sur toutes les variables mesurables. Il importe de noter que l'appariement reposant sur toutes les variables disponibles produirait généralement des estimations cohérentes, alors que le rajustement par les MCO pourrait ne pas accomplir ceci en raison de l'hypothèse de linéarité inhérente aux MCO.</p> <p>** L'emploi d'algorithmes exhaustifs d'appariement peut influencer l'efficacité des méthodes d'estimation VI.</p> <p>*** On tient pour acquis que les régressions de MCO comprendraient les trois types de variables mesurables, alors que l'appariement ne serait fondé que sur les ensembles de variables identifiées dans le tableau.</p>				

Pour expliquer ce tableau, nous donnerons deux exemples. Premièrement, supposons que les seules différences entre les participants et les membres du groupe témoin se retrouvent au chapitre d'éléments facilement mesurables si on dispose de données administratives complètes (c'est-à-dire qu'ils diffèrent uniquement selon des mesures V1 et V2). Dans ce cas, c'est la troisième colonne — la colonne « V1 ou V2 » — qui s'applique, et le tableau 1 montre que pratiquement toutes les procédures d'estimation donneraient d'assez bons résultats, peu importe comment les échantillons sont appariés¹⁸. Dans ce cas, la meilleure solution première consisterait à appairer à la fois V1 et V2, quoique la méthode des moyennes rajustées par régression et des procédures d'appariement relativement modestes

¹⁷ Ainsi, Heckman, Lalonde et Smith (1999) affirment que des estimateurs économétriques qui sont valides pour des échantillons aléatoires peuvent ne pas l'être lorsqu'ils sont appliqués à des échantillons établis par appariement (1939).

¹⁸ On tient pour acquis que les estimateurs V1 utilisent un variable instrumentale solide. Le recours à un instrument moins efficace pourrait produire des estimations incohérentes, même si la sélection à l'intérieur du programme n'est fondée que sur V1 et V2.

(V1) pourrait aussi fonctionner comme il faut. D'autre part, dans le cas plus réaliste où il existe entre les participants et les membres du groupe témoin des différences selon des dimensions non mesurées (« U »), la qualité des estimations fluctuerait. Ainsi, l'une des formules proposées à l'occasion conjugue l'appariement à partir de données administratives et des méthodes VI pour tenir compte des variables inobservables; c'est ce choix que l'on retrouve dans le coin inférieur droit du tableau 1. Le tableau nous permet de faire deux conclusions importantes au sujet de ce choix analytique :

- a. les propriétés d'une méthode reposant sur un appariement partiel et des procédures VI (Heckman) ne sont pas bien comprises. Le problème fondamental vient de ce que la cohérence des méthodes VI repose sur l'hypothèse voulant que le groupe témoin représente un échantillon aléatoire provenant d'une population plus vaste de participants éventuels. Un appariement partiel avant que ne soit estimée l'équation de la participation de premier degré utilisée dans la plupart des procédures VI serait inconciliable avec cette hypothèse. Pour déterminer quelle serait l'efficacité des procédures VI avec l'appariement partiel qu'on peut faire dans le contexte des PEMS, il faudrait faire d'autres recherches. En l'absence de résultats concluants à l'effet que cette approche produise des estimateurs cohérents avec une approche particulière, le groupe d'experts serait généralement opposé à ce genre d'approches mixtes;
- b. d'autres procédures — particulièrement celles qui reposent sur les écarts dans les différences — peuvent donner d'aussi bons résultats pour régler les problèmes attribuables à des différences non mesurées entre les participants et les membres du groupe témoin lorsque ces différences sont constantes dans le temps. Les procédures VI ne représentent donc pas nécessairement la meilleure stratégie dans tous les cas, même en supposant que des variables inobservables suscitent des différences importantes entre participants et membres du groupe témoin.

Ces énoncés ne s'appliquent évidemment qu'à un seul choix analytique possible. Tous les autres choix qui pourraient être proposés devraient faire l'objet d'une analyse semblable. Il est probable qu'un tel examen approfondi de diverses approches arrive à la conclusion que « tout dépend de la nature des variables inobservables ». C'est pourquoi le groupe d'experts pense que, dans la mesure du possible, il conviendrait d'examiner en détail différentes stratégies analytiques pendant la phase de conception de chaque évaluation et que *plusieurs* formules parmi les plus prometteuses devraient être retenues pour l'analyse à proprement parler.

3. Précision des estimations

Non seulement est-il souhaitable que les estimations des répercussions du programme établies dans le cadre des évaluations soient cohérentes, mais encore, pour être utiles, doivent-elles être relativement précises. En d'autres termes, l'éventail des répercussions vraies qui correspond aux estimations calculées dans le cadre d'une évaluation doit être relativement étroit. Si ce n'est pas le cas, l'éventail peut être tellement vaste qu'il sera impossible de tirer des conclusions au sujet de l'efficacité du programme — en fait, il se peut même que les données permettent de conclure de façon plausible que les

répercussions ont été nulles (voire même négatives). Il convient d'apporter quatre arguments généraux au sujet de la précision dans les évaluations. Premièrement, l'objectif de la cohérence doit, en toute logique, passer avant celui de la précision. Si une évaluation ne permet pas d'établir des estimations cohérentes, à quoi bon¹⁹ savoir que les estimations (inexactes) ont été mesurées précisément. Deuxièmement, la précision dépend souvent de la volonté de consacrer des ressources à une évaluation. En un sens, on peut toujours « acheter » plus de précision en accroissant simplement la taille des échantillons, de sorte qu'il y aura toujours un compromis entre la précision et la volonté de payer pour obtenir l'information qu'une évaluation est censée produire. Un troisième élément concerne la relation entre la précision et le choix des estimateurs. Certains des estimateurs figurant au tableau 1 nécessiteront des estimations reposant sur des sous-échantillons de données seulement²⁰, d'où une réduction inévitable de la précision. Un argument connexe plus général : pour un grand nombre des estimateurs énumérés au tableau 1, il est impossible de se prononcer clairement au sujet de la précision. En fait, dans certains cas, il est même impossible de calculer de façon analytique l'incertitude associée à une estimation et il faut l'établir de façon approximative à l'aide de méthodes de simulation d'auto-amorçage.

Compte tenu des complexités qui surgissent lorsqu'il s'agit de se prononcer sur la précision des estimations établies selon des méthodes non expérimentales, il est devenu courant d'utiliser des estimations de précision indicatives reposant sur des modèles expérimentaux dans la conception des évaluations. Ces indications tiennent implicitement pour acquis que les résultats obtenus par des méthodes expérimentales sont les plus précis qu'on puisse obtenir dans le cadre d'une évaluation, et les plans reposent sur ce scénario de la meilleure éventualité. Le tableau 2 donne un exemple de cette procédure. Il montre l'amplitude de la différence qui pourrait être décelée avec une efficacité de 80 p. 100, en supposant qu'un test de signification unilatéral à un seuil de 95 p. 100 soit utilisé dans les évaluations. Les écarts-types utilisés dans ces calculs proviennent de l'étude de faisabilité réalisée en Nouvelle-Écosse pour le projet indicateurs à moyen terme (IMT). Il se peut qu'ils sous-évaluent le niveau réel de précision qu'on pourrait obtenir lors des évaluations sommatives, car celles-ci pourraient compter sur de meilleures variables de contrôle et de meilleures méthodes statistiques que celles qui ont été utilisées pour l'étude de la Nouvelle-Écosse. Cependant, comme nous le verrons dans la section C, les estimations de la précision figurant dans le tableau correspondent approximativement aux constats de nombreuses évaluations expérimentales des programmes actifs du marché du travail.

¹⁹ En fait, dans certains cas, il pourrait être possible d'imposer des limites aux biais inhérents à une technique d'estimation en particulier, et on pourrait opter pour un compromis entre des effets estimés avec précision accompagnés de biais limités et des estimations cohérentes estimées de façon imprécise. Toutefois, en général, les causes des manques de cohérence ne sont pas suffisamment bien comprises pour qu'on puisse faire de tels compromis globaux.

²⁰ C'est particulièrement vrai des estimateurs d'appariement qui reposent sur des données d'enquête et suivant lesquels certains membres des groupes témoins qui ne sont pas appariés aux participants seraient éliminés de l'analyse.

TABLEAU 2				
Différences décelables, efficacité de 80 % et seuil de 95 %				
Taille des échantillons		Résultats obtenus par les participants (ET présumé)		
Participants	Témoins	Rémunération selon les T4 (11 000)	Occupent un emploi (0,370)	Semaines d'AE (9,00)
2 000	1 000	1 103	0,037	0,90
2 000	2 000	901	0,030	0,74
3 000	1 500	901	0,030	0,74
4 000	2 000	780	0,026	0,64
1 000	1 000	1 274	0,043	1,04
1 000	400	1 685	0,057	1,38
500	400	1 911	0,064	1,56

En tenant pour acquis que ces chiffres représentent raisonnablement les résultats qu'on peut attendre des évaluations, plusieurs observations peuvent être faites :

- 1. Il conviendrait de prévoir des nombres suffisants d'observations dans les groupes témoins au moment de la conception.** En effet, en comparant les deuxième et troisième rangées du tableau, par exemple, on constate que des échantillons équilibrés 2 000/2 000 donnent exactement le même degré de précision que des échantillons déséquilibrés 3 000/1 500. Aux fins de l'estimation des différences entre les résultats des participants et ceux des témoins, 500 personnes ont été interviewées en vain dans le second cas.²¹ Les deux dernières rangées témoignent également de l'importance que revêt la taille du groupe témoin. Elles illustrent les cas où l'appariement à l'aide de données d'enquête entraîne une réduction importante de la taille effective des échantillons. Il appert que dans de tels cas, si un programme a des répercussions modestes, elles ne seront vraisemblablement pas décelées à des seuils acceptables de signification et d'efficacité.
- 2. Les concepteurs des évaluations devraient examiner soigneusement s'il sera possible de déceler des différences suffisamment petites entre les témoins et les participants à des interventions particulières.** Les calculs d'efficacité montrent que seuls les effets très prononcés seront décelables, étant donné la taille prévue des échantillons aux fins de l'analyse de diverses interventions de prestations d'emploi(PE). Parallèlement aux problèmes éventuels qui peuvent se poser dans la sélection des groupes témoins pour des interventions données, ces constats nous amènent à penser que les attentes relatives à la mesure des répercussions

²¹ Ces plans déséquilibrés peuvent s'avérer utiles si le coût des cas de participants et le coût des cas témoins varient considérablement. Mais dans la plupart des évaluations non expérimentales, tel n'est pas le cas. Même au niveau des évaluations expérimentales, on remarque habituellement que l'observation des participants est plus coûteuse que celle des groupes témoins. Ainsi, des concepts déséquilibrés devraient inclure des groupes témoins proportionnellement plus gros.

d'interventions individuelles doivent être assez modestes, tout comme les possibilités d'estimer les répercussions pour d'autres sous-groupes ou à l'intérieur d'une province.²²

3. Il conviendrait d'envisager des échantillons plus vastes pour les évaluations.

Même si les chiffres du tableau indiquent que la plupart des différences entre participants et non-participants pouvant témoigner du rapport coût-efficacité de l'intervention devraient être décelables, compte tenu des tailles d'échantillon prévues (on envisage environ 3 000 entrevues), certains des chiffres, particulièrement ceux qui concernent les améliorations au chapitre de la rémunération, posent des problèmes. Les différences décelables pourraient facilement devenir déraisonnablement grandes si des préoccupations relatives à la validité donnent lieu à l'élimination de vastes portions de certains des groupes témoins. On devrait examiner dans quelle mesure des échantillons de plus grande taille permettraient d'atténuer ces problèmes.

4. Il conviendrait de faire un suivi attentif des taux de réponse dans les évaluations. Compte tenu des nombreux problèmes qui entourent la validité des approches analytiques dans les évaluations, il faudrait déployer tous les efforts possibles pour s'assurer que les données ne sont pas entachées de problèmes plus banals (dont certains seront examinés dans la prochaine section). Les données administratives disponibles devraient permettre de calculer l'ampleur de la non-réponse dans les évaluations et on devrait s'attendre à une analyse de cette question dans le cadre des évaluations.

²² Étant donné les échantillons beaucoup plus vastes inclus dans les données administratives qui doivent servir pour le projet des IMT, on peut penser que ce dernier est l'instrument tout indiqué pour l'examen des répercussions dans les sous-groupes.

C. Plan d'échantillonnage

Il faut répondre à trois grandes questions au moment de concevoir les échantillons qui seront utilisés dans les évaluations sommatives : 1) Comment le groupe des participants sera-t-il sélectionné? 2) Comment les données administratives et les données d'enquête seront-elles utilisées de concert pour la sélection du ou des groupes témoins? 3) Comment les ressources devraient-elles être réparties en fonction des échantillons?²³

1. Sélection des participants :

Selon le groupe d'experts, le tirage d'un échantillon de participants pour les évaluations ne pose pas de graves problèmes conceptuels (en supposant que les données administratives soient de qualité suffisante). Les participants seraient sélectionnés à partir des données administratives, selon la méthode qui représente le mieux l'activité du programme pendant une période donnée. Le groupe d'experts a fait trois recommandations mineures au sujet de ce processus de sélection :

- a. l'échantillon des participants devrait s'étaler sur une année complète²⁴, de façon à minimiser les influences saisonnières qui peuvent se manifester dans les résultats;
- b. il conviendrait de stratifier l'échantillon selon le genre de prestations d'emploi (PE) et la région. On assurerait ainsi une représentation adéquate des interventions comptant relativement peu de participants. Cette formule permettrait également de tenir compte des variations qu'on peut retrouver dans le contenu du programme d'une région à l'autre;
- c. l'échantillon devrait être tiré de façon à ce qu'il se soit écoulé un certain temps entre la fin du programme et le début de l'enquête. En termes concrets, au moins une année devrait s'être écoulée entre la fin de la période visée par l'échantillon et le début de la période d'enquête²⁵.

²³ On considère ici que les données administratives sont abondantes et qu'il n'en coûte rien de les recueillir. La taille des échantillons pendant la phase de collecte des données administratives des évaluations est donc considérée comme étant illimitée. Cependant, pour certaines interventions peu fréquentes, ce ne sera peut-être pas le cas, de sorte que nous discutons ici brièvement de la répartition de l'échantillon parmi les interventions.

²⁴ Si on retenait l'exercice financier, cela faciliterait également les comparaisons par rapport à d'autres données administratives — particulièrement si ce sont les dates de début qui sont utilisées pour définir la participation.

²⁵ Si les enquêtes étaient menées sur une période d'un an, deux années se seraient alors écoulées depuis la date de début du programme. Si les enquêtes sont regroupées pour des fins d'efficacité des entrevues, le groupe d'experts recommande un délai plus long entre la fin de la période visée par l'échantillon et le début des entrevues (peut-être 18 mois ou plus).

2. Sélection du groupe témoin

La sélection d'un groupe témoin est l'un des éléments les plus importants de la conception d'une évaluation. Pour simplifier l'examen de cette question, nous tenons pour acquis que toute l'analyse sera fondée sur des estimations moyennes des répercussions des prestations d'emploi et mesures de soutien (PEMS) sur les participants avec rajustement par régression. En d'autres termes, nous écartons pour l'instant certains des problèmes relatifs aux estimations que nous avons examinés dans la section B, pour nous concentrer explicitement sur les problèmes de sélection du groupe témoin. La question de savoir si les conclusions que nous présentons ici seraient différentes si d'autres stratégies d'estimation étaient utilisées (par exemple la méthode des écarts dans les différences ou des procédures VI) devra faire l'objet d'autres recherches, comme nous le verrons dans la section F.

On dispose de trois sources possibles de données pour la sélection du groupe témoin : 1) les données administratives concernant l'assurance-emploi (AE); 2) les données administratives de l'Agence des douanes et du revenu du Canada (ADRC) sur la rémunération et sur d'autres éléments relatifs à l'impôt; et 3) des données d'enquête. Il est important de comprendre les avantages et les inconvénients de ces diverses sources et comment ils peuvent influencer d'autres aspects de l'analyse.

a. Données concernant l'AE : Ces données sont les plus facilement accessibles pour la sélection du groupe témoin. Les témoins pourraient être choisis selon leurs antécédents en matière d'AE et/ou selon des données concernant des cessations d'emploi antérieures (les données des relevé d'emploi (RE)). Cette procédure donnerait probablement des résultats assez médiocres pour ce qui est de l'appariement par rapport aux antécédents des participants en matière d'emploi. Ce serait particulièrement vrai dans le cas des anciens prestataires réadmissibles de l'AE — ceux pour lesquels il n'y a pas actuellement de demande active de prestations. Même s'il y avait moyen de tirer un échantillon témoin de personnes ayant présenté des demandes de prestations dans le passé, celles-ci auraient vraisemblablement une expérience d'emploi plus récente que les anciens prestataires réadmissibles. Partant, même si l'appariement reposant uniquement sur les données de l'AE était considéré comme une méthode adéquate pour le groupe des prestataires actifs (ce qui n'a pas été démontré), il faudrait adopter des procédures supplémentaires pour les anciens prestataires réadmissibles.

b. Données sur la rémunération de l'ADRC : Les données sur la rémunération provenant de l'ADRC peuvent jouer un rôle essentiel dans la conception des évaluations sommatives. On sait²⁶ que les tendances de la rémunération pendant la période précédant immédiatement un programme sont un prédicteur important de la participation (Ashenfelter, 1978, Heckman et Smith, 1999). Plus généralement, on

²⁶ C'est du moins un constat fréquent dans les évaluations des programmes actifs du marché du travail aux États-Unis. Heckman, Lalonde et Smith (1999) suggèrent que ce phénomène de « baisse Ashenfelter » pourrait bien être mondial. L'ampleur du phénomène chez les participants aux PEMS n'est pas connue, mais certains travaux préliminaires portant sur des données recueillies en Nouvelle-Écosse (Nicholson, 2000) portent à croire que cette baisse survient aussi dans le cas de ce programme.

pense qu'une méthode permettant de tenir compte adéquatement des tendances de la rémunération serait un moyen prometteur pour régler les problèmes d'évaluation posés par les variables inobservables (Ashenfelter, 1979). Cette hypothèse est confirmée par certaines estimations préliminaires des répercussions des interventions des PEMS en Nouvelle-Écosse (Nicholson, 2000), qui montrent comment les données de l'ADRC peuvent être utilisées pour sélectionner un groupe témoin général de façon à ce qu'il ressemble davantage au groupe des participants²⁷. Malheureusement, on ne sait pas encore dans quelle mesure les évaluateurs des PEMS auront accès à ces données. Mais compte tenu de la méthode de sélection proposée et du calendrier de réalisation de l'enquête, il serait possible, selon les normes précédentes d'accessibilité, d'établir un profil exhaustif de la période antérieure au programme pour pratiquement tous les participants et les témoins éventuels. Peu importe qu'on choisisse une méthode générale pour sélectionner un groupe témoin ou qu'on ait recours à une méthode d'appariement par paires à titre individuel, il semble fort probable qu'on puisse obtenir une concordance assez rapprochée au chapitre de la rémunération observée.

- c. Données d'enquête :** Troisième source possible de données pour la sélection du groupe témoin : l'enquête de suivi qui serait menée environ deux ans après l'inscription au programme des PEMS. L'avantage de cette source, c'est qu'elle donne l'occasion de recueillir des données cohérentes et très détaillées auprès des participants et des témoins éventuels au sujet de leur activité sur le marché du travail avant le programme, et d'autres données ayant trait à l'inscription éventuelle à des interventions PEMS. Ces données peuvent être utilisées dans le cadre de différentes méthodes (y compris l'appariement selon les caractéristiques et la propension, et certaines procédures VI) permettant de vérifier les différences au niveau des participants/comparaisons.

Les avantages éventuels des données d'enquête au moment de définir les méthodes analytiques qui seront utilisées dans le cadre des évaluations ne devraient cependant pas faire oublier les inconvénients de ces données, et notamment :

- Des données d'enquête sur les activités préalables au programme ne représenteront pas une base de référence véritable. On recueillera plutôt les données à l'aide de questions qui demanderont aux répondants de se souvenir d'événements qui ont eu lieu il y a plusieurs années. Les **erreurs de mémoire** dans de telles enquêtes peuvent être très importantes — et elles se répercuteront directement sur les méthodes qui comptent beaucoup sur ce genre de données d'enquête rétrospectives.
- L'utilisation de données d'enquête pour sélectionner les groupes témoins se traduira nécessairement par une certaine **réduction de la taille des échantillons** qui serviront en fin de compte à l'analyse — tout simplement parce qu'il pourra

²⁷ D'autres variables provenant des données fiscales de l'ADRC pourraient servir à l'appariement des participants et des témoins, par exemple le revenu total, le revenu familial total, le nombre de personnes à charge, etc.

arriver que certains répondants ne fassent pas l'affaire comme membres du groupe témoin. L'ampleur de cette réduction dépendra beaucoup de la portée de l'appariement qui peut être fait à partir des données administratives. En l'absence de données provenant de l'ADRC, ces réductions pourraient se révéler considérables. Ainsi, une grande partie des fonds consacrés à l'enquête pourra, en fin de compte, avoir été dépensée sans servir à l'analyse.

- Finalement, il se peut que l'utilisation de données d'enquête pour la sélection des groupes témoins **compromette les données primaires sur les résultats** qui doivent être recueillies dans le cadre de l'enquête. De toute évidence, ce problème se produirait s'il faut consacrer tellement de temps à recueillir des informations exhaustives relatives à la période d'avant le programme dans le cadre de l'enquête que la collecte de données plus détaillées relatives à la période d'après le programme risque d'en souffrir. En outre, et plus subtilement, si on recueille des données exhaustives visant la période d'avant le programme et la période d'après le programme dans le cadre d'une même enquête, il se peut que les répondants adaptent leurs réponses de telle façon que des biais latents s'y manifestent. Même si cette possibilité ne repose pas sur des preuves empiriques solides, le personnel qui effectue l'enquête vérifiera souvent par en avant et par en arrière un ensemble de réponses spécifiques pour voir s'il y a cohérence. Ceci peut améliorer la qualité des données mais également biaiser la déclaration des données rétrospectives et, par conséquent, certaines estimations d'impact.

3. Approches proposées

S'agissant du problème de la spécification des groupes témoins, on peut envisager deux démarches :

- a. L'approche idéale :** Il semble clair que, compte tenu des données qu'il y aurait moyen d'obtenir aux fins des évaluations, l'approche idéale pour la sélection du groupe témoin serait d'utiliser à la fois des données de l'AE et des données de l'ADRC pour constituer un échantillon témoin qui se rapproche étroitement de l'échantillon des participants, selon des critères observables dans ces deux sources de données. L'appariement par paires pourrait être possible au moyen de ces vastes fichiers de données administratives, mais ne sera peut-être pas la meilleure méthode dans une perspective globale, selon les stratégies analytiques précises qui doivent être mises en œuvre. Par exemple, une stratégie axée davantage sur la sélection d'un échantillon à des fins d'appariement pourrait se révéler préférable dans les applications où on se propose d'utiliser des procédures d'estimation VI, parce que la sélection peut retenir plus de la variance sous-jacente au niveau des variables qui déterminent la participation. L'appariement par paires pose également certains problèmes logistiques de mise en œuvre à l'étape de l'enquête. On pourrait faire des enquêtes auprès d'un **échantillon aléatoire de paires**, mais la non-réponse de la part de l'un ou l'autre des membres d'une paire aggraverait considérablement le biais de non-réponse. Bien sûr, aucun mode d'appariement fondé sur des variables observables ne peut garantir que des différences non observables n'entraîneront pas,

en bout de ligne, des résultats incohérents. Il reste qu'une approche ayant recours à des données de l'AE et de l'ADRC pour la sélection du groupe témoin semblerait la plus prometteuse dans bien des cas. Par contre, il serait nettement avantageux de pouvoir recourir à certaines recherches sur la meilleure façon d'utiliser l'appariement de première étape au niveau d'une évaluation qui reposerait, au bout du compte, sur des données d'enquête.

b. L'approche de rechange : Si on n'a pas accès aux données de l'ADRC à des fins de sélection de l'échantillon dans le cadre d'une évaluation, il faudrait se rabattre sur diverses procédures qui sont nettement de second choix. On commencerait par un appariement approximatif à partir des données accessibles de l'AE. Les sources de données et les variables à utiliser pourraient varier selon les évaluations, d'après la nature des marchés du travail locaux et les participants aux interventions PE. Après cet appariement approximatif, toutes les autres procédures relatives aux groupes témoins reposeraient sur les données d'enquête. Cette approche aurait trois grandes conséquences pour la conception globale des évaluations :

- i. **l'enquête devrait être plus longue**, pour permettre de recueillir une information adéquate sur les activités au sein du marché du travail avant le programme;
- ii. **il faudrait un groupe témoin plus nombreux** par rapport à l'approche « idéale » (voir la sous-section suivante), pour tenir compte de la présence possible de sujets qui finiraient par se révéler non comparables;
- iii. **le rôle relatif des méthodes d'appariement devrait être réduit** dans les évaluations (ne serait-ce qu'en raison de la réduction de la taille des échantillons qui est inévitable lorsqu'on a recours à une méthode d'appariement reposant sur des données d'enquête) et le rôle des procédures VI²⁸ devrait être élargi. En raison de la nécessité de recourir aux procédures VI, il peut s'avérer nécessaire de consacrer des ressources additionnelles au développement de « bons » instruments, soit en procédant à la collecte de données additionnelles²⁹ ou en engageant un processus quelconque de répartition au hasard.

4. Taille des échantillons

Vu les difficultés et les incertitudes liées à la conception des évaluations sommatives, il est difficile de se prononcer définitivement sur la taille souhaitable des échantillons. Le groupe d'experts est néanmoins convaincu qu'à ce sujet, on peut tirer certaines conclusions d'autres évaluations — particulièrement celles qui ont fait appel à la

²⁸ On pourrait également utiliser de façon plus exhaustive les méthodes des écarts dans les différences, bien que leur application à des données provenant d'une seule enquête ouvre la possibilité de corrélations dans les erreurs de déclaration qui risquent de biaiser les résultats au fil du temps.

²⁹ La cueillette de données sur une étendue géographique pour les fournisseurs de services, par exemple, pourrait s'avérer un instrument utile dans le cas de certains emplacements.

répartition aléatoire. Comme, en principe, les échantillons répartis de façon aléatoire ne posent pas de problèmes analytiques particuliers, on peut les considérer comme des « scénarios de référence » à partir desquels les modèles non expérimentaux peuvent être jugés. Dans des conditions idéales, où tous les postulats se vérifient, un modèle non expérimental serait l'équivalent d'une expérience avec répartition au hasard, une fois appliquées les méthodes analytiques appropriées (par exemple, l'appariement par paires ou des techniques VI). Ainsi, les expériences avec répartition aléatoire qui ont déjà été menées constituent un modèle intéressant pour les évaluations.

Le tableau 3 indique la taille des échantillons utilisés pour l'analyse³⁰ de quelques-unes des grandes évaluations à répartition aléatoire³¹ menées aux États-Unis.

TABLEAU 3			
Taille des échantillons de certaines études à répartition aléatoire			
Évaluation	Taille de l'échantillon expérimental	Taille de l'échantillon témoin	Nombre de traitements
National JTPA	13 000	7 000	3
Illinois UI Bonus	4 186	3 963	1
NJ UI Bonus	7 200	2 400	3
PA UI Bonus	10 700	3 400	6
WAUI Bonus	12 000	3 000	6
WA Job Search	7 200	2 300	3
SC Claimant	4 500	1 500	3
Supported Work	3 200	3 400	1
S-D Income Maint.	2 400	1 700	7
National H.I.	2 600	1 900	3

Diverses tendances ressortent de ce tableau sommaire :

- les échantillons sont tous d'assez grande taille — les groupes témoins comportent au moins 1 500 personnes et, plus généralement, 2 000 personnes ou plus;
- les expériences qui prévoient un seul traitement optent généralement pour une répartition égale entre le groupe expérimental et le groupe témoin³²;
- les évaluations comportant de multiples traitements attribuent des portions relativement plus vastes de leurs échantillons aux catégories de traitement; toutefois, les groupes témoins sont généralement plus vastes que n'importe quelle cellule à traitement unique;

³⁰ La taille finale de ces échantillons tient compte de la non-réponse à l'enquête ou à certaines de ses questions. Il conviendrait d'accroître la taille des échantillons initiaux pour tenir compte de cette érosion.

³¹ Pour obtenir un résumé de plusieurs de ces évaluations, ainsi qu'un grand nombre de références, voir Greenberg et Shroder (1997).

³² Ce qui minimise la variance de l'effet estimé de traitement pour un budget d'évaluation donné, en supposant que les cas de traitement et les cas témoins sont tout aussi coûteux.

- même si ce n'est pas apparent dans le tableau, un grand nombre des évaluations ont eu recours à une méthode de traitement « étagée », qui permettait de créer des traitements plus complexes en ajoutant des composantes à des traitements simples (ce qui a été le cas de bien des évaluations concernant l'assurance-chômage, par exemple). Dans ce cas, les traitements simples peuvent faire office de « contrôles » pour les traitements plus complexes, car la formule permet de mesurer les effets différentiels des nouveaux traitements³³. Ainsi, le nombre réel de « contrôles » peut être sous-déclaré dans le tableau pour ces évaluations.

Comme un grand nombre de ces évaluations cherchaient à mesurer des résultats assez semblables à ceux qui doivent être mesurés dans les évaluations des PEMS, les tailles de ces échantillons sembleraient pertinentes dans le cas des PEMS. Plus précisément, on peut conclure de ces expériences que pour être efficaces, les groupes témoins doivent se composer d'au moins 2 000 personnes³⁴. C'est la nature des traitements à examiner dans l'évaluation des PEMS qui justifie des groupes témoins aussi vastes en termes relatifs. Comme les cinq grandes interventions offertes en vertu du programme des PEMS dans les régions sont assez différentes les unes des autres, il ne sera pas possible de réaliser des gains d'efficacité grâce aux modèles étagés caractéristiques des expériences relatives à l'AC³⁵. L'hétérogénéité des caractéristiques des participants aux interventions PEMS individuelles est une autre raison de recourir à un vaste groupe témoin. Lorsqu'il s'agit d'évaluer des interventions à titre individuel, il est probable que seule une petite partie du groupe témoin puisse être utilisée pour chaque intervention, ce qui milite également en faveur d'un vaste groupe témoin.

L'évaluation d'interventions PE particulières soulève d'autres complications au moment de la répartition des échantillons. Dans la plupart des régions, ce sont les interventions DC qui sont de loin les plus courantes. Une méthode simple de sélection au hasard des interventions risquerait donc de se traduire par des échantillons de taille très réduite pour les autres interventions. C'est pourquoi il est probable que la plupart des évaluations recourent à des échantillons de participants stratifiés selon l'intervention et accordent plus de poids aux interventions plus restreintes³⁶. Si une évaluation recourt à cette structure, le choix d'une structure de comparaison devient complexe. Le groupe témoin devrait-il lui aussi être stratifié de façon à ce que chaque sous-groupe corresponde aux participants à

³³ Dans un grand nombre des évaluations, toutefois, les traitements moins élaborés se révèlent souvent les plus efficaces. C'est le cas de pratiquement toutes les évaluations expérimentales ayant trait à l'AC.

³⁴ Cette conclusion est grosso modo conforme aux calculs indicatifs de l'efficacité de la section B (qui eux-mêmes reposent sur les variances observées dans les données de la Nouvelle-Écosse). Précisons à nouveau que les tailles suggérées ici sont celles des échantillons définitifs. En d'autres termes, il s'agirait du nombre de cas pouvant effectivement être analysés, une fois éliminés les cas inadmissibles, la non-réponse aux enquêtes, et ainsi de suite. De plus, lorsqu'il faut recueillir des données sur le revenu, il serait bon que les enquêtes sur le terrain soient menées à l'époque des déclarations d'impôt, pour que les répondants aient ces renseignements sous la main et en mémoire pendant l'entrevue.

³⁵ Toutefois, on pourrait envisager un modèle étagé en intégrant une cellule concernant les SAE seulement à certaines évaluations. Les expériences relatives à l'assurance-chômage aux États-Unis laissent penser qu'un traitement concernant exclusivement les SAE pourrait effectivement présenter certains effets décelables.

³⁶ Plus généralement, l'échantillon des participants pourrait être réparti entre les interventions de référence en fonction des intérêts des décideurs régionaux. Le plan d'attribution pourrait donc représenter un compromis entre la représentation proportionnelle et la stratification en cellules de taille égale.

telle ou telle strate d'intervention? Ou devrait-il être établi en fonction des caractéristiques d'un échantillon aléatoire de participants? Cette dernière formule pourrait être préférable si les procédures VI devaient jouer un rôle de premier plan dans l'analyse, puisqu'un groupe de comparaison plus large pourrait aider à prévoir l'affectation aux traitements. D'autre part, le recours à des groupes de comparaison propres à l'intervention pourrait s'avérer préférable si les procédures d'estimation sont plus simples. On devrait s'attendre à ce que les évaluateurs examinent soigneusement ces enjeux lorsqu'ils conçoivent les évaluations.

5. Questions relatives à la mise en œuvre

La mise en œuvre de toute stratégie d'appariement dans le cadre d'une évaluation qui compte sur des données d'enquête pose des problèmes particuliers. L'objectif général est d'éviter de « gaspiller » des entrevues auprès de membres du groupe témoin qui ne seront jamais utilisées ultérieurement dans l'analyse. Pour cela, il faut généralement qu'une bonne partie des entrevues auprès des participants ait été réalisée avant les entrevues auprès des témoins. De cette façon, les caractéristiques des membres du groupe témoin peuvent être rajustées en fonction de l'échantillon des participants³⁷. Une stratégie de sondage encore plus complexe pourrait se révéler nécessaire si on ne dispose que de données administratives très limitées pour la sélection de l'échantillon. En pareil cas, il pourra se révéler nécessaire de prévoir certaines questions d'élimination dans les questionnaires des membres du groupe témoin, de sorte qu'on puisse mettre fin aux entrevues des cas non comparables avant d'y avoir consacré trop de temps. Cette procédure a été suggérée, par exemple, pour régler le problème de la sélection d'un groupe témoin pour les participants du groupe des anciens prestataires réadmissibles. Comme de telles procédures de sélection d'un échantillon reposant sur une enquête peuvent comporter des biais, il convient de faire preuve d'une extrême prudence si on les adopte.

³⁷ L'adaptation de procédures de ce genre pour régler le problème de la non-réponse peut être très délicate, particulièrement si les taux de non-réponse ne sont pas les mêmes dans le groupe des participants et le groupe témoin. C'est une autre raison qui explique pourquoi les enjeux relatifs à la non-réponse méritent une place de premier plan dans la conception des évaluations.

D. Spécification et mesure des résultats obtenus par les participants

Les réflexions du groupe d'experts au sujet de la spécification et de la mesure des résultats reposent sur quatre critères : 1) il semble probable que pour la plupart des évaluations, certaines des principales mesures³⁸ porteront sur l'emploi après le programme³⁹; 2) il faudrait donc recueillir suffisamment d'information relative à l'emploi pour que diverses mesures puissent être établies — ce qui facilitera la conception des mesures des résultats en fonction des objectifs particuliers des programmes⁴⁰; 3) il conviendrait de recueillir des données sur diverses autres variables socioéconomiques importantes — qui serviront surtout de variables de contrôle dans l'étude des résultats en matière d'emploi; et 4) les techniques utilisées pour la collecte des données d'enquête devraient viser une comparabilité directe entre diverses évaluations régionales.

Voici sept mesures concrètes conformes à ces critères :

- 1. Sauf considérations particulières, l'enquête de suivi devrait avoir lieu au moins deux ans après la date de début du plan d'action.** De la sorte, il serait plus probable que le plan d'action soit achevé bien avant la date de l'entrevue. Comme, selon les premières évaluations, les entrevues devraient avoir lieu à l'automne 2001, elles devraient viser les plans d'action dont la date de début tombe au cours de l'exercice 1999 (entre le 1er avril 1998 et le 31 mars 1999). Les évaluations dont les entrevues auront lieu plus tard pourraient se concentrer plutôt sur l'exercice 2000.
- 2. Lorsque c'est possible, il conviendrait d'utiliser dans toutes les évaluations des questions semblables sur les antécédents des répondants en matière d'emploi.** Comme c'est l'emploi après le début de l'intervention et après le programme qui fera probablement l'objet de la plupart des évaluations des PEMS, il semble clair qu'une bonne partie des ressources des enquêtes devrait être consacrée à sa mesure. Des études antérieures ont montré que l'utilisation d'instruments différents pour la collecte des données peut mener à des résultats estimatifs assez différents (Heckman, Lalonde et Smith, 1999). Pour éviter cette source de variance, il conviendrait d'encourager les évaluateurs à utiliser les mêmes batteries de questions. Dans la même vaine, les routines d'épuration des données et de

³⁸ L'achèvement du programme, les prestations d'assurance-emploi (AE) touchées et l'inscription subséquente à d'autres interventions des prestations d'emploi et mesure de soutien (PEMS) pourraient constituer d'autres mesures.

³⁹ En réalité, l'emploi devrait généralement être mesuré à partir de la date d'inscription au programme (et d'une date semblable pour les groupes témoins) puisque cela procure l'évaluation la plus précise des occasions d'emploi abandonnées relativement à la participation au programme.

⁴⁰ En réalité, l'emploi peut être mesuré à partir de la date d'inscription au programme si elle se révèle le meilleur point de référence pour l'échantillon des participants et pour celui du groupe témoin.

construction des variables devraient être coordonnées d'une évaluation à l'autre. Les évaluateurs devraient également envisager la meilleure façon d'exploiter la documentation appartenant aux répondants (par exemple, les documents fiscaux) pour la collecte de ces données.

3. Il conviendrait d'élaborer un certain nombre de mesures de l'emploi. Il s'agit ici de veiller à ce que les mesures des résultats utilisées dans les évaluations correspondent bien aux objectifs prévus du programme. Même si l'on peut s'attendre à ce que toutes les évaluations élaborent les mêmes mesures de base (par exemple, le nombre de semaines d'emploi au cours de l'année écoulée, la rémunération globale pendant cette période, le nombre d'emplois occupés, etc.), il y aurait par ailleurs une certaine flexibilité pour que les régions puissent se concentrer sur les mesures qu'elles jugent les plus appropriées, compte tenu de l'éventail des prestations offertes. Par exemple, pour les clients des interventions de la composante Développement des compétences, les résultats pourraient se concentrer sur les augmentations salariales après le programme ou sur les changements dans le genre des emplois occupés. Pour les participants du volet Subventions salariales ciblées, les évaluations pourraient se concentrer sur les transitions réussies vers des emplois non subventionnés. Et il peut se révéler très difficile de concevoir des moyens de mesurer la viabilité à long terme des options de travail indépendant choisies par certains clients. Comme on l'a vu précédemment, il faudra de toute évidence faire d'autres recherches pour déterminer précisément les liens qui seront établis entre les résultats et les interventions. À un niveau plus conceptuel, il faudra montrer explicitement comment les résultats qui doivent être mesurés dans le cadre des évaluations correspondent aux objectifs sociétaux généraux du programme des PEMS (par exemple, tels qu'ils sont énoncés dans la loi habilitante).

4. Il conviendrait d'élaborer un module de base pour la collecte de données sur d'autres variables socioéconomiques qui pourrait être utilisé dans la plupart des évaluations. Il s'agirait ici de s'accorder sur les variables intermédiaires à mesurer et de veiller à ce qu'elles soient examinées dans toutes les évaluations. Faute d'un tel accord, il pourra être très difficile de comparer les résultats des analyses d'une région à l'autre, parce que les analyses porteraient sur différents facteurs dans différentes évaluations. La mise en commun des données à des fins d'analyse entre les régions serait également entravée si les évaluations ne comptaient pas ce module commun. De toute évidence, de telles analyses comparatives entre les régions présentent un intérêt direct pour Développement des ressources humaines canada (DRHC), parce qu'elles permettront à la fois d'accroître la précision de certaines estimations et de faire ressortir certains déterminants importants des divers degrés de réussite entre les régions⁴¹. Par conséquent, il faudrait envisager des moyens d'encourager tous les évaluateurs à utiliser des modules de base semblables — peut-être en chargeant un entrepreneur de les élaborer en vertu d'un contrat distinct.

⁴¹ On reviendra sur l'intérêt de mettre en commun les données de différentes évaluations dans la section F.

- 5. Il conviendrait d'envisager d'autres entrevues de suivi, au moins dans certaines régions.** Même si la plupart des évaluations auront probablement recours à la formule des enquêtes uniques, le groupe d'experts est convaincu qu'il conviendrait d'encourager les évaluateurs à examiner ce que des suivis ultérieurs pourraient révéler (peut-être 24 mois après l'enquête initiale). Il est probable que la collecte de ces données supplémentaires soit particulièrement utile dans les cas où les interventions ne promettent des résultats qu'à relativement long terme. Les entrevues supplémentaires, si elles sont jugées cruciales pour une évaluation, seraient probablement confiées à des entrepreneurs indépendants. Peu importe qu'une entrevue de suivi soit prévue ou non dans le cadre de l'évaluation, le groupe d'experts estime que DRHC devrait prendre des dispositions pour que les participants à l'évaluation puissent être suivis dans le temps, à partir des données administratives relatives à l'AE et (en principe) à la rémunération (voir l'argument suivant).
- 6. Il conviendrait d'utiliser des données administratives dans toutes les évaluations pour mesurer certains résultats obtenus par les participants.** Il se peut que des questions de temps empêchent l'utilisation de données administratives sur la rémunération (tirées des feuillets T4) pour mesurer les résultats obtenus par les participants dans le cadre des évaluations telles qu'elles ont été données à contrat actuellement (bien qu'on puisse envisager des contrats de suivi à cet égard), mais il conviendrait d'exploiter le plus possible les données administratives de l'AE. Celles-ci peuvent fournir les mesures les plus exactes des prestations d'assurance-emploi touchées par les participants et elles peuvent également nous éclairer sur la validité des résultats de l'enquête en ce qui concerne l'emploi. On peut également utiliser des données administratives dans le cadre des évaluations pour élaborer des mesures semblables à celles qui ont été élaborées pour le projet des indicateurs à moyen terme (IMT), ce qui faciliterait la comparaison entre les deux études (voir la section E ci-dessous). L'utilisation de données administratives pour mesurer les résultats obtenus par les participants présente également des avantages qui iraient bien au-delà des contrats d'évaluation à titre individuel. En principe, il devrait être possible de suivre les membres du groupe des participants et du groupe témoin pendant de nombreuses années grâce à ces données. Leur utilisation permettrait de déterminer si les répercussions du programme observées lors des évaluations se sont maintenues ou ont disparu rapidement. Il est également possible que les fichiers longitudinaux plus longs que permettraient d'établir les données administratives nous renseignent sur la validité des estimations initiales des répercussions, car ils exploiteraient plus pleinement les variations mesurées dans les caractéristiques de la rémunération des participants et des membres du groupe témoin au fil du temps.
- 7. Il conviendrait d'envisager des analyses coûts-avantages et des analyses de rentabilité, mais elles joueront vraisemblablement un rôle secondaire dans les évaluations.** Les estimations de l'impact établies dans le cadre des évaluations pourraient jouer un rôle important en servant de base à des analyses coûts-avantages et à des analyses de rentabilité. Il serait facile de faire une analyse de rentabilité relativement simple, en supposant qu'on dispose de données au sujet des coûts

différentiels des interventions. L'utilité d'une telle analyse dépend beaucoup de la capacité d'estimer fidèlement les répercussions d'interventions particulières — autant du point de vue des modèles quasi expérimentaux que de celui de la détermination de la taille adéquate des échantillons si on veut obtenir des estimations suffisamment détaillées. Il pourrait quand même être possible de faire certaines comparaisons rudimentaires entre les interventions.

La réalisation d'analyses coûts-avantages exhaustives à partir des évaluations pourrait cependant poser des difficultés plus grandes, notamment en raison de l'importance des budgets prévus. Entre autres obstacles majeurs susceptibles de nuire à la tenue d'une analyse coûts-avantages exhaustive, soulignons le fait qu'un grand nombre de répercussions sociales découlant du programme des PEMS pourraient être difficiles à mesurer, qu'il peut s'avérer difficile de prévoir l'impact à long terme des programmes et que, étant donné la taille globale du programme, les effets de déplacement pourraient être significatifs dans certaines régions. Il est particulièrement difficile de trouver des moyens de régler ce dernier problème de façon simple. Pour toutes ces raisons, le groupe d'experts est convaincu que les budgets modestes qui ont été prévus pour les diverses évaluations ne permettraient pas de réaliser les recherches qui s'imposeraient pour faire une analyse coûts-avantages viable. Mais il croit fermement qu'une sorte quelconque d'analyse coûts-avantages devrait constituer le but ultime des évaluations, car les parties intéressées voudront savoir si la valeur des programmes est conforme à leurs coûts. Le groupe croit, par conséquent, qu'il pourrait être prudent que DRHC examine des moyens de mener séparément une analyse coûts-avantages à partir des données combinées⁴² de diverses évaluations.

⁴² À tous le moins, il y aurait lieu d'encourager les fournisseurs à fournir des ensembles de données pour usage public que d'autres chercheurs pourraient combiner pour mettre au point des analyses coûts-avantages ou d'autres types d'analyses.

E. Coordination avec le projet des indicateurs à moyen terme (IMT)

Les évaluations des ententes sur le développement du marché du travail (EDMT) joueront un rôle important dans le processus qui mènera à l'élaboration d'indicateurs à moyen terme pour le programme des PEMS (le « projet des IMT »). Comme ces indicateurs seront élaborés principalement à partir de données administratives, les évaluations donnent l'occasion de déterminer les lacunes éventuelles d'indicateurs fondés seulement sur des données relativement limitées. Une telle comparaison pourra contribuer à déterminer si le projet des IMT réussit à saisir tous les effets (positifs et négatifs) des interventions de prestations d'emploi (PE). Les données plus riches recueillies dans le cadre des évaluations peuvent également représenter des moyens simples d'améliorer les indicateurs à moyen terme prévus (par exemple, si les données de plusieurs sources administratives sont combinées ou si des méthodes novatrices sont utilisées pour regrouper les données au fil du temps). Par ailleurs, le projet des IMT, étant donné que ses échantillons et ses activités récurrentes seront beaucoup plus vastes, permettra d'examiner les effets sur divers sous-groupes et les innovations récentes apportées au programme, qu'il serait impossible d'examiner dans le cadre des évaluations de EDMT. Les effets des IMT dans le cadre de telles analyses seront plus utiles si les évaluations ont permis d'établir des mesures de validité de base montrant les rapports entre ces effets estimés et les mesures plus détaillées des résultats provenant des évaluations de EDMT. Il sera donc essentiel que les deux projets soient coordonnés. Aux fins de cette coordination, le groupe d'experts a formulé les recommandations suivantes :

- 1. Les échantillons des évaluations devraient être sélectionnés de façon à faciliter la comparaison avec les résultats du projet des IMT.** Un grand nombre des suggestions faites dans les sections A et B ci-dessus visent cet objectif. En général, on espère que les échantillons des évaluations pourront être considérés comme des échantillons aléatoires des populations plus vastes qui seraient utilisées pour l'élaboration des IMT pendant une période donnée. De cette façon, les résultats des évaluations permettraient de faire des analyses directes des IMT. Il faudra tenir des consultations pendant les phases d'élaboration des évaluations des EDMT pour veiller à ce que cet objectif soit respecté.
- 2. Les entrepreneurs chargés des évaluations de EDMT devraient, lorsque c'est possible, élaborer des mesures semblables aux IMT pour leurs échantillons (ou faciliter la tâche d'autres chercheurs qui s'en chargeront).** Pour examiner l'efficacité des indicateurs à moyen terme et la façon dont ils pourraient être améliorés, il faudra utiliser des microdonnées tirées des évaluations. Cela ne sera toutefois possible que si l'on s'est donné la peine d'intégrer les données administratives appropriées aux fichiers de recherche. De toute évidence, certaines données qui serviront à l'élaboration des IMT ne seront peut-être pas accessibles aux évaluateurs dans les délais prévus (ce sera probablement le cas des données plus

récentes de l'Agence des douanes et du revenu du Canada (ADRC). Dans ces circonstances, les fichiers de recherche devraient être conçus de telle façon que ces données puissent y être intégrées ultérieurement.

- 3. On devrait s'efforcer de coordonner les évaluations prévues et le processus d'élaboration des IMT.** On risque de rater un grand nombre des possibilités de coordination entre les évaluations des EDMT et le projet des IMT si les deux projets ne sont pas menés approximativement en même temps. Puisque les dates de réalisation des évaluations sont plus ou moins déterminées en vertu des EDMT, les décideurs qui s'intéressent à l'élaboration des IMT devraient, lorsque c'est possible, s'inspirer du même calendrier.

F. Sommaire des questions sur lesquelles il faudra revenir au moment de la mise en œuvre

Ce document présente un cadre générique de pointe pour les évaluations des prestations d'emploi et mesures de soutien (PEMS). Au moment de mettre en œuvre les méthodes qui y sont exposées, les réalisateurs des évaluations devraient tenir compte de la situation particulière dans la région visée. Ils pourront ainsi justifier toute hypothèse reposant sur le cadre, de même que les modifications et les extensions. Celles-ci devraient être bien étayées et rationalisées en fonction des constats, des raisons qui justifient l'application des techniques, de même que des connaissances et des théories sous-jacentes. Les résultats seront d'autant plus crédibles que plusieurs techniques auront été utilisées simultanément, selon la méthode des recoupements. De surcroît, pour tenir compte de certaines des incertitudes inhérentes aux modèles quasi expérimentaux, il conviendrait d'encourager les réalisateurs à fournir des fourchettes d'estimations, plutôt que des estimations monovalentes.

On a abordé ici de façon très générale certains des grands enjeux qui se manifesteront probablement au fur et à mesure de la conception des évaluations des PEMS. Notre objectif n'était pas de fournir un plan explicite pour chacune de ces évaluations, mais plutôt de décrire certaines des questions méthodologiques universelles sur lesquelles il faudra se pencher. Le groupe d'experts est convaincu que les évaluations des PEMS donnent l'occasion de faire des travaux de recherche de pointe en matière d'évaluation, tout en recueillant une information substantielle qui sera à la fois valable et utile pour les décideurs.

Le groupe d'experts a déterminé diverses questions sur lesquelles il faudra se pencher de plus près. C'est aux réalisateurs qu'il reviendra de s'occuper des détails.

- 1. La répartition des tâches entre les diverses évaluations devrait se révéler profitable.** Comme on ne peut pas s'attendre à ce que toutes les évaluations fassent « tout », le groupe d'experts pense qu'une certaine répartition des tâches entre les évaluations serait appropriée et qu'elle permettrait d'obtenir les meilleurs résultats globaux. Voici trois exemples de répartition des tâches : 1) une analyse plus approfondie de certaines interventions (p. ex., le travail indépendant) dans certaines évaluations; 2) certaines évaluations pourraient mettre davantage l'accent sur les résultats de non emploi plus difficiles à mesurer; 3) si une évaluation donne l'occasion de mettre en œuvre une répartition au hasard, il conviendrait de l'exploiter.
- 2. Il conviendrait d'élaborer des procédures permettant de déterminer la qualité des résultats des évaluations et de les faire connaître.** Même si certaines personnes qui ne font pas partie de la collectivité des chercheurs seront convaincues que *les conclusions* des évaluations des PEMS ne comporteront aucune surprise pour elles,

il semble beaucoup plus probable que les résultats réels soient assez différents d'une région à l'autre. Une partie de cette variation pourra être attribuable à des différences véritables dans les répercussions du programme, une autre partie à des différences dans la conception et la mise en œuvre du programme, une autre encore à des différences dans la conception des évaluations, et à des différences dans la demande de main-d'œuvre à l'échelle des régions et enfin, une autre partie à des facteurs purement aléatoires (p. ex., des différences au chapitre du marché du travail). Le groupe d'experts pense que si la communication de ces variations ne se fait pas de façon ordonnée, c'est la validité globale perçue des évaluations qui risque d'en souffrir. Pour atténuer de tels problèmes, il semblerait souhaitable que d'entrée de jeu, on mette sur pied des procédures de communication des résultats des évaluations selon lesquelles ces résultats seraient replacés dans un contexte global. Ce contexte exposerait les aspects particuliers de chaque évaluation à titre individuel, tout en expliquant comment ces résultats correspondent à un profil plus global.

3. Il conviendrait d'envisager des moyens de mettre en commun les résultats des évaluations dans les analyses subséquentes. Le groupe d'experts pense que la mise en commun des résultats des évaluations, même si elle peut poser des problèmes de non-comparabilité statistique, devrait faire partie intégrante du processus d'évaluation des PEMS, et que les fournisseurs devraient être encouragés à rendre leurs données disponibles à de telles fins. On y retrouverait divers avantages, notamment les suivants :

- des échantillons plus vastes, de sorte que la précision des estimations s'en trouverait améliorée;
- la capacité d'étudier les effets de différences éventuelles dans le contenu des programmes;
- des indications sur le rendement des interventions de prestations d'emploi (PE) dans différents marchés du travail;
- de meilleures possibilités d'examiner les résultats obtenus par certains sous-groupes de participants;
- la possibilité d'explorer des enjeux qui vont peut-être au-delà du mandat des évaluations à titre individuel — par exemple entreprendre une analyse de rentabilité détaillée du programme dans son ensemble.

Pour atteindre ces objectifs importants, il sera nécessaire d'imposer un certain degré de comparabilité entre les évaluations des points de vue de la collecte et de l'accessibilité des données. Des normes en ce sens doivent être prévues dès le début — il se révélera peut-être impossible de mettre en commun les données de plusieurs évaluations si chaque évaluateur a toute latitude pour choisir ses propres solutions aux problèmes concernant les données.

4. Il conviendrait d'envisager, lorsque c'est possible, d'utiliser des instruments communs de collecte des données dans toutes les évaluations. Cette conclusion dérive des trois arguments précédents. Les problèmes relatifs à la mise en commun et à la communication des résultats des évaluations seront tous exacerbés si chaque évaluation a ses propres instruments de collecte des données. Le groupe d'experts est donc fermement convaincu que certains modules communs (particulièrement en ce qui concerne les antécédents sur le marché du travail et les résultats obtenus par les participants) devraient être intégrés à toutes les évaluations. Pour élaborer de tels modules, il faudra prévoir des consultations exhaustives entre divers intervenants du programme des PEMS, à la fois pour veiller à ce que l'instrument de collecte des données soit de qualité élevée et pour faire en sorte que des données soient recueillies sur tous les résultats jugés pertinents du point de vue des politiques publiques (y compris divers résultats socioéconomiques qui n'ont pas été examinés explicitement ici).

5. Quel degré d'appariement pourra-t-on faire avant le tirage des échantillons? Comme on l'a vu dans les sections B et C, il est impossible de concevoir un plan d'échantillonnage pour les évaluations tant qu'on ne connaît pas le degré d'appariement qui pourra être fait à partir des données administratives. C'est pourquoi le groupe d'experts est convaincu qu'il est important d'entreprendre certains travaux préliminaires à partir des données administratives, pour déterminer les limites du processus d'appariement. En général, on peut s'attendre à ce que ces limites soient plus importantes s'il est impossible d'utiliser les données de l'agence des douanes et du revenu du Canada (ADRC) pour le tirage des échantillons dans les premières évaluations. Il sera essentiel que les lacunes que représentera l'utilisation des seules données de l'AE pour choisir les groupes témoins soient étayées au moment de l'application de la méthodologie et que des moyens de combler ces lacunes soient proposés dans le cadre de l'enquête. Il sera particulièrement important de déterminer dans quelle mesure l'incapacité d'établir de bons appariements à partir des seules données de l'AE influencera la répartition de l'échantillon entre les groupes de participants et les groupes témoins.

6. Quel rôle les estimations VI devraient-elles jouer dans les évaluations? L'utilisation d'estimateurs VI soulève certaines des questions les plus difficiles en ce qui concerne les évaluations. En effet, la détermination des situations dans lesquelles de tels estimateurs donneraient des estimations cohérentes des répercussions est souvent ambiguë; en outre, la détermination de la qualité des estimateurs relève autant de l'art que de la science. Le groupe d'experts croit donc fermement que le rôle des estimateurs VI dans les évaluations devrait être examiné. Voici quelques-unes des questions auxquelles il conviendrait de s'intéresser en ce qui concerne l'utilisation de tels estimateurs :

- Quelles variables seront utilisées aux fins de l'identification que nécessitent les estimateurs VI? Comment ces données seront-elles recueillies? Comment les restrictions de l'identification seront-elles testées?

- Comment les propriétés des estimateurs VI seront-elles influencées par la façon dont les échantillons des participants et des témoins seront sélectionnés? Une analyse de simulation peut-elle permettre de comprendre la relation entre l'estimation VI et l'appariement de l'échantillon?
- Comment la robustesse des estimateurs VI peut-elle être évaluée? Comment devrait-on procéder pour déclarer et évaluer les autres spécifications possibles pour les estimateurs VI?

7. Il conviendrait d'examiner plus à fond la question de la taille des échantillons pour chaque application. Selon les calculs présentés ici, on aurait besoin d'échantillons d'au moins 2 000 participants et 2 000 témoins pour obtenir que les estimations seront suffisamment précises pour déterminer les impacts pertinents des politiques. Mais cette estimation de précision s'est faite à partir d'une information très limitée ou uniquement à partir de comparaisons par rapport à d'autres études. De toute évidence, un examen plus minutieux de la question propre à chaque évaluation s'impose avant que d'importantes ressources soient engagées. Un tel examen devrait porter sur des questions comme les suivantes : 1) le désir d'estimer les impacts d'interventions particulières; 2) comment le groupe témoin peut être « partagé » entre différentes interventions, s'il peut l'être du tout; 3) comment les méthodes d'analyse se répercuteront vraisemblablement sur la taille des échantillons disponibles; et 4) comment la probabilité de non-réponse influence le calcul de la taille des échantillons.

Bibliographie

ASHENFELTER, Orley. « Estimating the Effects of Training Programs on Earnings », *Review of Economics and Statistics*, janvier 1978, p. 47-57.

ASHENFELTER, Orley. « Estimating the Effect of Training Programs on Earnings with Longitudinal Data », dans *Evaluating Manpower Training Programs*, ouvrage collectif publié sous la direction de F. Bloch, Greenwich, CT, JAI Press, 1979, p. 97-117.

ASHENFELTER, Orley, et David CARD. « Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs », *Review of Economics and Statistics*, juin 1985, p. 648-660.

DEHEJIA, Rajeev, et Sadek WAHBA. « Propensity Score Matching Methods for Nonexperimental Causal Studies », National Bureau of Economic Research, Cambridge, MA, Working Paper No. 6829, 1998.

DEHEJIA, Rajeev, et Sadek WAHBA. « Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs », *Journal of the American Statistical Association*, décembre 1999, 94(448), p. 1053-62.

GREENBERG, David, et Mark Shroder. « The Digest of Social Experiments, Second Edition », Washington, D.C., The Urban Institute Press, 1997.

HECKMAN, James. « Varieties of Selection Bias », *American Economic Review Papers and Proceedings*, mai 1990, 80(2), p. 313-18.

HECKMAN, James, et Joseph HOTZ. « Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training », *Journal of the American Statistical Association*, décembre 1989, 84(408), p. 862-74.

HECKMAN, James, Hidehiko ICHIMURA, Jeffrey SMITH et Petra TODD. « Characterizing Selection Bias Using Experimental Data », *Econometrica*, septembre 1998a, 66(5), p. 1017-98.

HECKMAN, James, Hidehiko ICHIMURA et Petra TODD. « Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme », *Review of Economic Studies*, octobre 1997, 64(4), p. 605-54.

HECKMAN, James, Hidehiko ICHIMURA et Petra TODD. « Matching as an Econometric Evaluation Estimator », *Review of Economic Studies*, avril 1998b, 65(2), p. 261-94.

HECKMAN, James, Robert LALONDE et Jeffrey SMITH. « The Economics and Econometrics of Active Labor Market Programs », dans *Handbook of labor economics*, vol. 3A, ouvrage collectif publié sous la direction de Orley Ashenfelter et David Card, Amsterdam, North-Holland, 1999, p. 1865-2097.

HECKMAN, James, et Jeffrey SMITH. « The Preprogramme Earnings Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies », *Economic Journal*, juillet 1999, 109(457), p. 313-48.

ICHIMURA, Hidehiko, et Christopher TABER. « Direct Estimation of Policy Impacts », National Bureau of Economic Research, Cambridge, MA, Technical Working Paper No. 254, 2000.

IMBENS, Guido, et Joshua ANGRIST. « Identification and Estimation of Local Average Treatment Effects », *Econometrica*, mars 1994, 62(2), p. 467-75.

LALONDE, Robert. « Evaluating the Econometric Evaluations of Training Programs with Experimental Data », *American Economic Review*, septembre 1986, 76(4), p. 604-20.

NEWHEY, Whitney, et James POWELL. « Instrumental Variables for Nonparametric Models », manuscrit inédit, Université Princeton, 1989.

NICHOLSON, Walter. « Assessing the Feasibility of Measuring Medium Term Net Impacts of the EBSM Program in Nova Scotia », document de travail rédigé pour DRHC, mars 2000.

NICHOLSON, Walter. « Problèmes de conception des évaluations sommatives : Sommaire des discussions... », DRHC, mars 2001.

ROSENBAUM, Paul, et Donald RUBIN. « The Central Role of the Propensity Score in Observational Studies for Causal Effects », *Biometrika*, avril 1983, 70(1), p. 41-55.

SMITH, Jeffrey, et Petra TODD. « Reconciling Conflicting Evidence on Performance of Propensity Score Matching Methods », *American Economic Review Papers and Proceedings*, mai 2001, 91(2), p. 112-18.

Annexe A⁴³ :

***Évaluations sommatives des prestations
d'emploi et mesures de soutien ainsi que
des ententes sur le développement du
marché du travail — répercussions
éventuelles des programmes à
examiner/mesurer***

La présente annexe expose un éventail de répercussions éventuelles des programmes. Cette « démarche des répercussions éventuelles des programmes » est conforme au cadre qu'utilise depuis quelque temps le Bureau du vérificateur général pour examiner la mesure de l'efficacité des programmes menée par les ministères et organismes fédéraux.

Aux fins de l'évaluation, il *ne s'agit pas* de mesurer les résultats individuels par rapport à un objectif quantitatif ou à un niveau-repère quantitatif précis à atteindre. La démarche consiste plutôt à examiner et à mesurer, de la manière la plus complète possible, ce que les programmes financés dans le cadre des (PEMS)/(EDMT) ont accompli (leurs activités) et ce qu'ont été les résultats.

Il s'agit ici d'adopter la même démarche d'évaluation de base dans toutes les régions, puis d'élaborer une méthode uniforme à l'échelle nationale pour l'analyse agrégée des résultats globaux des PEMS/EDMT. Par ailleurs, une marge de manœuvre sera ménagée au niveau local en ce qui concerne les thèmes qui seront abordés dans les évaluations; ainsi, les régions pourront décider d'y intégrer d'autres questions/indicateurs, à leur choix.

Il est présumé que les consultants retenus rédigeront, une fois le processus d'évaluation amorcé, des rapports de méthodologie abordant plus à fond le processus de mesure dans les secteurs choisis.

⁴³ La présente annexe ne fait pas partie du cadre élaboré par le groupe d'experts. Leurs débats ont porté en grande partie sur la méthodologie expliquée dans le rapport. Les discussions au sujet d'éléments bien précis à mesurer ont été très limitées. La clarification des problèmes de mesures pourrait constituer une lourde tâche future pour les membres du groupe d'experts.

La section intitulée « Répercussions éventuelles du programme... », plus bas, énumère des questions considérées importantes dans plusieurs politiques liées à l'emploi et dans des études menées par différents chercheurs et gouvernements, à l'échelle nationale ou internationale. En ce qui concerne les évaluations des PEMS, les questions peuvent être tirées des éléments suivants:

- les objectifs des EDMT/PEMS tels qu'on les trouve dans la législation sur l'assurance-emploi (AE) et les ententes relatives au marché du travail;
- les questions soulevées dans le cadre d'évaluations, d'analyses de politique ou bien d'énoncés de politique en rapport avec l'emploi;
- les questions soulevées et examinées dans le cadre des évaluations formatives des EDMT/PEMS;
- les questions soulevées lors de la planification préliminaire des évaluations sommatives et dans le cadre du projet-pilote des indicateurs à moyen terme;
- les répercussions éventuelles que des investissements de cette nature peuvent avoir sur le marché du travail et l'économie en général;
- les tendances socioéconomiques émergentes au Canada et chez ses partenaires commerciaux.

Comme il en est question dans le rapport, la méthodologie comprend diverses techniques de mesure qui peuvent être appliquées à un vaste éventail de résultats ou d'indicateurs. Par conséquent, le cas échéant, Développement des ressources humaines Canada (DRHC) s'attendra à ce que les résultats soient mesurés comme suit : résultats obtenus par les participants après le programme, différences entre les expériences avant et après le programme, et comparaisons des expériences des participants et des non-participants.

RÉPERCUSSIONS ÉVENTUELLES DU PROGRAMME RÉSULTATS OBTENUS PAR LES PARTICIPANTS

CLIENTS QUI ONT REÇU DE L'AIDE À TITRE INDIVIDUEL

- POURCENTAGE DES PARTICIPANTS QUI ONT TROUVÉ DES EMPLOIS DANS LA PÉRIODE QUI A SUIVI LE PROGRAMME
- POURCENTAGE DES SEMAINES DE TRAVAIL DANS LA PÉRIODE QUI A SUIVI LE PROGRAMME
- DURÉE MOYENNE DES PÉRIODES D'EMPLOI
- NIVEAUX MOYENS DES PRESTATIONS D'AE/D'AIDE SOCIALE TOUCHÉES
- POURCENTAGE DES SEMAINES DE PRESTATIONS D'AE/D'AIDE SOCIALE DANS LA PÉRIODE QUI A SUIVI LE PROGRAMME (autosuffisance des travailleurs — réduction de la dépendance envers l'AE. et l'aide sociale)

- RECOURS À L'AE PAR RAPPORT À L'AIDE SOCIALE AVANT ET APRÈS LE PROGRAMME
- RÉMUNÉRATION MOYENNE PAR RAPPORT AU SEUIL DE FAIBLE REVENU (ou autres mesures de la rémunération ou du revenu utilisées sur le marché ou au sein de la collectivité)
- REVENU MOYEN PAR RAPPORT AU SEUIL DE FAIBLE REVENU (ou autres mesures de la rémunération ou du revenu utilisées sur le marché ou au sein de la collectivité)
- INDICATEURS DE LA QUALITÉ DE VIE (ET SURTOUT LES AMÉLIORATIONS QUE CERTAINS PARTICIPANTS AFFICHENT PAR RAPPORT À CEUX QUI N'ONT CONNU AUCUN CHANGEMENT : MOTIVATION; SATISFACTION FACE À LA VIE). (Nota : c'est pour les raisons suivantes que cette question est intégrée à la démarche proposée pour l'évaluation : a) parce qu'il s'agit de l'un des objectifs globaux officiels du ministère; et b) parce que dans les ouvrages techniques publiés sur la question en général [y compris les ouvrages d'autres pays], de tels effets peuvent être associés à des programmes du genre de ceux qui sont prévus dans les EDMT)
- MOBILITÉ GÉOGRAPHIQUE/RÉINSTALLATION DES CLIENTS

PRINCIPALES CARACTÉRISTIQUES DE LA CLIENTÈLE (et surtout les principaux genres de problèmes auxquels font face les clients qui reçoivent de l'aide en vertu des EDMT).

- ÂGE/SEXE PAR RAPPORT À L'ENSEMBLE DES CHÔMEURS (CHÔMAGE CHEZ LES JEUNES / ENJEUX DE L'ÉQUITÉ ENTRE LES HOMMES ET LES FEMMES)
- PROPORTION DES CLIENTS QUI ÉPROUVENT DES PROBLÈMES CHRONIQUES D'EMPLOI/QUI SONT AUX PRISES AVEC DES OBSTACLES À L'EMPLOI AVANT LA PARTICIPATION AU PROGRAMME (d'après une recension de la littérature sur les principaux genres d'obstacles à l'emploi)
- PROPORTION DES CLIENTS DONT ON A DÉTERMINÉ QU'ILS ÉTAIENT DES TRAVAILLEURS DÉPLACÉS AVANT DE RECEVOIR L'AIDE DU PROGRAMME
- NIVEAUX DE SCOLARITÉ (ENJEUX DE L'ALPHABÉTISATION FONCTIONNELLE)

- CATÉGORIES PROFESSIONNELLES ET NIVEAUX DE QUALIFICATION AVANT ET APRÈS LE PROGRAMME (lacunes dans les compétences/acquisition de compétences)
- APPRENTIS (MATÉRIEL DESCRIPTIF SEULEMENT)

COLLECTIVITÉS LOCALES

- INITIATIVES DE DÉVELOPPEMENT DU MARCHÉ DU TRAVAIL POUR AIDER LES COLLECTIVITÉS (PAR. EX., AIDE AUX ORGANISMES LOCAUX DE DÉVELOPPEMENT ÉCONOMIQUE/EMPLACEMENT DES INDUSTRIES) (MATÉRIEL DESCRIPTIF SEULEMENT)

GOUVERNEMENTS

- RAPPORT COÛT-EFFICACITÉ DES INTERVENTIONS (par ex., améliorations nettes de l'employabilité et de la durée d'emploi parmi diverses catégories de clients par rapport aux coûts de l'aide fournie dans le cadre du programme)
- LES PEMS/EDMT COMME MÉCANISME D'OFFRE DE MAIN-D'ŒUVRE (évaluation de l'ampleur des activités mises en œuvres dans le cadre des EDMT par rapport aux besoins de l'économie en matière d'expansion et de remplacement des effectifs)
- ÉCONOMIES BRUTES POUR LE COMPTE DE L'AE DÉCOULANT DES PEMS/EDMT
- ÉCONOMIES NETTES POUR LE COMPTE DE L'AE DÉCOULANT DES PEMS/EDMT

SUIVI DES ÉTUDES FORMATIVES

- CERTAINES QUESTIONS SOULEVÉES DANS LE CADRE DES ÉTUDES FORMATIVES
- CERTAINES QUESTIONS QUI ONT ÉMERGÉ DEPUIS LA FIN DES ÉTUDES FORMATIVES