

N° 95-19 au catalogue

**COMBINAISON DES DONNÉES ADMINISTRATIVES ET
DES DONNÉES D'ENQUÊTE EN VUE D'ALLÉGER LE
FARDEAU DES RÉPONDANTS DANS LES ENQUÊTES
LONGITUDINALES**

Numéro d'enregistrement du produit 75F0002M

Août 1995

Sylvie Michaud, Division des méthodes d'enquêtes sociales

David Dolson, Division des méthodes d'enquêtes sociales

Donna Adams, Division des enquêtes-ménages

Martin Renaud, Division des méthodes d'enquêtes sociales

La série de documents de recherche de l'EDTR est conçue en vue de communiquer les résultats des études ainsi que les décisions importantes ayant trait à l'Enquête sur la dynamique du travail et du revenu. Ils sont offerts en français et en anglais. Pour obtenir une description sommaire des documents disponibles ou un exemplaire de ces documents, communiquez avec Philip Giles, EDTR, par la poste à Édifice Jean-Talon, 11^e étage, section D8, Statistique Canada, Ottawa (Ontario), Canada, K1A 0T6; par INTERNET: GILES@STATCAN.CA; par téléphone au (613) 951-2891; ou par télécopieur au (613) 951-3253.

SOMMAIRE

Le présent article a été présenté aux assises annuelles de 1995 de l'American Statistical Association à Orlando, en Floride.

Depuis de nombreuses années, l'usage des dossiers administratifs aux fins de statistiques va croissant. En règle générale, on recourt à ces dossiers parce qu'ils constituent pratiquement la seule source d'information pour d'importantes statistiques. Ils complètent souvent les données recueillies lors des enquêtes, sans toutefois les remplacer. Cet article explique de quelle manière il est possible d'alléger le fardeau des répondants en substituant les données administratives aux données d'enquête dans certains cas, mais non tous les cas.

Un des objectifs de l'Enquête sur la dynamique du travail et du revenu de Statistique Canada consiste à mesurer l'incidence que les changements qui affectent le marché du travail et la situation familiale auront sur le revenu des particuliers. Chaque panel de l'enquête longitudinale est d'une durée de six ans, et ses membres répondent à deux interviews par année. La première, en janvier, a pour but de recueillir des renseignements sur le travail effectué l'année précédente tandis que la seconde, en mai, sert à obtenir de l'information sur le revenu. Les questions sur le revenu sont les plus exigeantes et par le fait même, plus sujettes à des problèmes au niveau de la qualité des données. En raison de ces préoccupations, les participants peuvent à leur gré répondre à un questionnaire ou autoriser Statistique Canada à consulter leur déclaration de revenus pour en extraire directement l'information. Les résultats de ces deux méthodes de collecte des données sont regroupés en un seul fichier de sortie. Le présent article décrit les effets prévus de cette méthode mixte sur les réponses et la qualité des données, et examine les problèmes de mesure connexes.

TABLE DES MATIÈRES

	Page
1. Introduction	1
2. Données administratives	1
3. L'Enquête sur la dynamique du travail et du revenu	4
4. Pourquoi recourir aux données fiscales dans le cadre de l'EDTR ?	6
4.1 Méthode classique	6
4.2 Interview assistée par ordinateur	6
4.3 Faisabilité de la méthode fiscale	10
4.4 Approche mixte	12
5. Implications et problèmes	14
6. Évaluation	17
6.1 Incidence potentielle sur le taux de réponse	17
6.2 Qualité des données détaillées	22
7. Conclusions	29
Bibliographie	32

1. INTRODUCTION

La demande de données et d'information des organismes de statistique nationaux ne cesse d'augmenter depuis de très nombreuses années. À cela s'est récemment ajoutée l'influence générale de budgets qui s'amenuisent constamment. On s'est toujours inquiété du fardeau des répondants mais l'importance de ce facteur s'est encore accrue au cours des dernières années, comme l'indiquent les preuves anecdotiques sur la réticence des personnes interrogées à diverses occasions.

L'an dernier, Statistique Canada a lancé trois nouvelles enquêtes longitudinales pour satisfaire de nouvelles demandes d'information : l'Enquête longitudinale nationale sur les enfants, l'Enquête nationale sur la santé de la population et l'Enquête sur la dynamique du travail et du revenu (EDTR). Les auteurs de politiques, les gestionnaires de programmes, les analystes et d'autres utilisateurs en retireront beaucoup. Néanmoins, le coût de telles enquêtes, tant sur le plan monétaire qu'à l'égard du fardeau imposé aux répondants, est loin d'être négligeable. En effet, les trois enquêtes portent sur un échantillon passablement important, le questionnaire est long et les répondants seront sollicités plusieurs fois pendant quelques années. Par conséquent, toutes les façons de réaliser des économies ou d'alléger le fardeau des répondants présentent une utilité. On se penchera ici sur une de ces enquêtes, soit l'Enquête sur la dynamique du travail et du revenu, dont les objectifs ont pu être atteints grâce à l'usage des données administratives, dans le cadre d'une méthode de collecte mixte.

2. DONNÉES ADMINISTRATIVES

Dans cette section, on parlera brièvement des données administratives, de leurs usages, de leurs avantages et de leurs inconvénients. L'information présentée s'inspire largement d'un document rédigé par G. Brackstone (1987).

Selon Brackstone, il existe six grandes catégories de dossiers administratifs, c'est-à-dire :

- 1) ceux qui servent à régulariser le passage des marchandises et des personnes à la frontière (on peut penser notamment aux dossiers sur les exportations et sur l'immigration);
- 2) ceux qui résultent d'exigences juridiques voulant que certains événements soient enregistrés (naissances, décès et constitutions de sociétés);
- 3) ceux nécessaires à l'administration des obligations et des avantages sociaux comme l'impôt et l'assurance-chômage;
- 4) ceux dont on a besoin pour diriger des institutions publiques (écoles, hôpitaux et tribunaux);
- 5) ceux issus de la réglementation de la branche d'activité par le gouvernement (les dossiers sur les banques et les sociétés de transport en sont un exemple);
- 6) ceux qui résultent de la prestation de services publics comme l'électricité et le téléphone.

Statistique Canada recourt aux dossiers des six catégories; dans le cas de l'EDTR, ce sont les déclarations de revenus dont on se sert.

L'évaluation des données d'enquête ou des données du recensement est l'un des premiers usages des dossiers administratifs. Ainsi, on a souvent fait appel aux déclarations de revenus et aux registres de l'immigration pour évaluer les données du Recensement de la population sur le revenu et l'immigration.

Les données administratives servent aussi à créer et à maintenir des bases de sondage. Par exemple, Statistique Canada se sert des dossiers des compagnies de téléphone pour établir la base de données qui servira aux enquêtes téléphoniques à composition aléatoire. Les renseignements sur les retenues à la source que les

employeurs envoient à Revenu Canada sont eux aussi couramment utilisés. Ces données servent en partie à maintenir le registre des entreprises puisqu'elles permettent d'identifier les nouvelles entreprises ou les modifications subies par les sociétés existantes.

Quelques grands programmes de statistique nécessitent la tabulation directe des données extraites des dossiers administratifs. C'est notamment le cas des statistiques sur le commerce extérieur et l'état civil.

Néanmoins, l'usage le plus fréquent des dossiers administratifs demeure sans doute l'estimation indirecte, les données administratives représentant un ou plusieurs des paramètres de l'estimation. Ainsi, on peut obtenir une estimation partielle des migrations en couplant les déclarations de revenus des particuliers d'une année à l'autre. Chaque année, Statistique Canada produit des estimations démographiques et économiques sur les particuliers et les familles à partir des déclarations de revenus. Au fil des ans, les dossiers pertinents ont été combinés pour former une base de données administratives longitudinale qui facilite l'analyse longitudinale des données fiscales se rapportant aux particuliers et aux familles. En regroupant les données d'enquête sur les grandes entreprises à l'information extraite des déclarations de revenus des petites entreprises, après correction si nécessaire, il est possible d'obtenir certaines statistiques sur la branche d'activité. Dans le cadre de l'Enquête sur l'emploi, la rémunération et les heures de travail, on a récemment commencé à se servir des données sur les retenues à la source au lieu des données d'enquête sur les petites entreprises. On a ainsi réalisé des économies appréciables et allégé sensiblement le fardeau des répondants, tout en améliorant la qualité des statistiques finales. L'usage des déclarations de revenus par l'EDTR aura des effets analogues.

Les sources de données administratives peuvent se traduire par d'importants avantages en ce qui concerne le coût de l'enquête, le fardeau des répondants et la qualité des données. Elles peuvent aussi faciliter les estimations relatives aux petites régions, à un coût modeste ou moyen. Toutefois, plusieurs facteurs déterminent l'utilité d'une source de données administratives pour un programme statistique. Les trois principaux sont les suivants. Dans quelle mesure les définitions et les principes du régime administratif épousent-ils ceux du programme statistique ? Quelle est la couverture du régime administratif et dans quelle mesure correspond-elle à celle nécessaire au programme ? Quelle sera la qualité des données tirées des dossiers administratifs ? Deux autres préoccupations sont la fréquence et l'actualité. Enfin, il ne faut pas non plus négliger la question de la stabilité. En effet, les régimes administratifs se trouvent rarement sous le contrôle des statisticiens. Ils peuvent donc subir des changements auxquels les statisticiens devront s'ajuster. Tous ces aspects doivent être réglés si on veut utiliser les déclarations de revenus dans le cadre de l'EDTR.

3. L'ENQUÊTE SUR LA DYNAMIQUE DU TRAVAIL ET DU REVENU

L'Enquête sur la dynamique du travail et du revenu est une nouvelle enquête par panel, lancée au Canada en 1994. Elle est conçue pour mesurer l'évolution de la situation économique dans le temps et fournir des renseignements sur les paramètres des changements observés, surtout en ce qui concerne la population, la famille et le marché du travail. L'enquête met l'accent sur la dynamique à moyen terme, le but étant de suivre les particuliers pendant six ans. Les répondants seront interviewés deux fois par année. La première interview, en janvier, servira à recueillir des renseignements sur le travail pour la dernière année civile complète. Une autre, en mai, permettra d'obtenir des renseignements sur le revenu de l'année précédente également. Les mois d'avril ou de mai constituent le moment idéal

pour recueillir des renseignements sur le revenu car à ce moment, le répondant aura sans doute sa déclaration de revenus à portée de la main. En la consultant, il pourra donner des renseignements plus précis sur son revenu.

Au lieu d'obtenir les renseignements sur le revenu par une interview, on pourrait effectuer un micro-appariement en reliant directement les données sur le travail recueillies à l'interview précédente au fichier du ministère du Revenu. Pareille approche pourrait améliorer la qualité des données et alléger le fardeau des répondants (qui subiraient six interviews au lieu de douze). Pour recourir à une telle approche cependant, il faut d'abord obtenir l'autorisation du répondant. (La *Loi sur la statistique*, à laquelle le Bureau doit se conformer, exige qu'on obtienne le consentement éclairé du répondant avant d'entreprendre pareilles activités de couplage des enregistrements.) Les résultats initiaux indiquent fortement qu'on ne pourrait abandonner totalement l'interview sur le revenu du mois de mai, car trop peu de répondants permettraient la consultation de leur déclaration de revenus. C'est pourquoi l'EDTR a adopté une approche «mixte» pour la collecte des données. Les répondants pourront à leur gré autoriser la consultation de leur déclaration de revenus ou répondre au questionnaire de l'enquête.

La méthode mixte ne va pas sans inconvénients. Le présent document examine les difficultés qu'il faut surmonter, décrit les études entreprises pour évaluer l'impact de cette approche particulière et présente les futurs projets de recherche en guise de conclusion.

4. POURQUOI RECOURIR AUX DONNÉES FISCALES DANS LE CADRE DE L'EDTR ?

4.1 MÉTHODE CLASSIQUE

La méthode classique adoptée par Statistique Canada pour recueillir les données sur le revenu des particuliers consistait à demander à ces derniers d'indiquer le revenu total de toute source réalisé l'année antérieure. L'Enquête sur les finances des consommateurs (EFC) annuelle demeure le principal outil pour cela. On a récemment commencé à recourir à l'interview assistée par ordinateur dans le cadre de cet exercice. Auparavant, on procédait à une interview avec papier et crayon, les répondants recevant le questionnaire et un guide explicatif avant l'appel de l'intervieweur. Pour certaines questions, le guide se rapportait aux lignes numérotées de la déclaration de revenus.

Cette méthode de collecte des données présente plusieurs lacunes. Ainsi, quelques sources de revenu sont sous-déclarées. Les revenus de placements, l'aide sociale, les prestations d'assurance-chômage et le revenu d'un travail autonome en sont des exemples importants. Après comparaison d'autres sources de données, Poulin (1993) a constaté qu'on sous-déclare ces revenus à raison de 10 à 50 % en général, l'importance de l'écart variant avec le type de revenu. Dans le contexte d'une enquête longitudinale comme l'EDTR, les erreurs de mesure sensibles au niveau des enregistrements revêtent une importance particulière, car une bonne partie de l'analyse s'effectue au niveau micro-économique.

4.2 INTERVIEW ASSISTÉE PAR ORDINATEUR

En 1993, on a commencé à évaluer une méthode permettant de rapprocher davantage la déclaration de revenus et les données sur le revenu de l'EDTR.

Poulin (1993) évoque les problèmes, les avantages, les inconvénients et les difficultés théoriques associés à cette approche. En mai 1993, on a testé une méthode qui favorisait la consultation de la déclaration de revenus de 1992 et encouragé les répondants à l'adopter - le questionnaire citait les lignes numérotées précises de la déclaration de revenus, et les répondants étaient priés d'inscrire les sommes apparaissant sur leur déclaration dans un «carnet» de l'EDTR.

Le but principal de l'exercice était surtout de réduire les erreurs, mais on espérait aussi qu'en simplifiant la collecte des données par référence directe à la déclaration de revenus, le taux de réponse serait plus élevé.

Néanmoins, quelques inconvénients possibles ont été cernés. Ainsi, le régime fiscal ne définit pas certaines composantes du revenu de la même façon que l'EFC et la plupart des enquêtes-ménages. Les sources de revenu n'étant pas toutes imposables, on doit recueillir des données sur certaines autrement qu'avec la déclaration de revenus. Parallèlement, on devrait pouvoir répondre au questionnaire de l'EDTR avec ou sans référence à la déclaration de revenus, car certains participants pourraient ne pas l'avoir sous la main, ni même en avoir remplie une. Enfin, la complexité du questionnaire de l'EDTR doit rester raisonnable, même s'il existe plusieurs types de déclaration de revenus pour les particuliers et si la plupart d'entre elles sont assez complexes.

En mai 1993, on a procédé à un essai auprès de 1500 ménages choisis dans deux provinces. Les répondants sélectionnés avaient participé à l'EFC l'année précédente et été interviewés en janvier 1993 lors de la collecte des données de l'EDTR sur le travail. À l'époque, l'EFC reposait encore sur l'interview avec papier et crayon (P et C). L'essai avait deux objectifs principaux : tester la méthode envisagée dans le cadre de l'EDTR pour recueillir les données sur le

revenu et déterminer la meilleure façon d'obtenir l'information sur le revenu au moyen de l'interview assistée par ordinateur (IAO).

Trois mécanismes de collecte des données ont été développés dans le cadre de l'EDTR, ainsi que Giles (1993) les décrit en détails. Avant l'essai, on a demandé à chaque répondant de remplir le «carnet» de questions qui leur avait été envoyé. Dans la mesure du possible, les références à la déclaration de revenus y étaient indiquées. L'usage du carnet rempli abrégait l'interview, si le répondant l'avait rempli, puisqu'il suffisait de lire les montants notés.

Pour faciliter la tâche des participants qui n'avaient pas rempli le carnet, l'ensemble complet de questions a été divisé en groupes ou «blocs» présentant un lien logique. Une question générale au début de chaque bloc servait à déterminer si la source de revenu spécifique s'appliquait. Le cas échéant, des questions plus précises suivaient.

On a également encouragé les répondants qui avaient leur déclaration de revenus à l'utiliser en citant les lignes numérotées du formulaire. Selon le type de déclaration, certains blocs de questions étaient automatiquement éliminés.

En l'absence de déclaration de revenus, on a recouru à l'approche des «blocs». Celle-ci s'inspire de la déclaration de revenus, mais aucun bloc n'a été escamoté et on n'a fait aucune référence aux lignes numérotées de la déclaration.

Outre les trois méthodes utilisées pour recueillir les données sur le revenu, deux autres grands aspects distinguent l'IAO de la méthode P et C. En effet, la première autorise une correction interactive. Ensuite, on introduit le processus d'interview dépendante grâce à une série d'indicateurs tirés de l'interview de janvier sur le travail. Ces indicateurs ont été utilisés lors de l'interview sur le

revenu pour obtenir des renseignements sur certaines composantes du revenu, quand elles n'étaient pas mentionnées. Par exemple, un travailleur rémunéré devait signaler un montant quelconque au titre des salaires et traitements.

On a évalué la qualité des données en les appariant aux données fiscales. Dans la mesure du possible, on a établi un lien direct grâce aux variables suivantes : nom, sexe, état matrimonial, âge, code postal, date de naissance et nom du conjoint. La concordance était parfaite dans environ 50 % des cas. Quand tel n'était pas le cas, on s'est servi du système généralisé d'appariement des enregistrements V1 de Statistique Canada (aussi appelé CANLINK) pour obtenir un appariement probabiliste. On a ainsi obtenu un taux d'appariement global de 84 %. Environ la moitié des répondants non appariés avaient signalé un revenu nul à l'EFC (pour l'année antérieure). Il semblait donc probable que bon nombre d'entre eux ne faisaient pas partie du fichier des déclarations de revenus utilisé pour l'appariement.

Grondin et Michaud (1994) ont rédigé un article présentant les résultats de cette étude. En règle générale, les données d'enquête et les données fiscales concordent plus avec l'IAO qu'avec la méthode P et C. Le taux de sous-déclaration est aussi plus faible avec l'IAO. Des trois mécanismes, celui du carnet et celui de la déclaration de revenus donnent clairement de meilleurs résultats que celui des blocs de questions. Le processus d'interview dépendante s'avère également efficace. La comparaison avec les données fiscales révèle que la présence d'un indicateur permet de prévoir correctement l'existence d'un montant quelconque.

On a décidé de ne recourir qu'à une approche - similaire à celle du carnet - lors de l'exercice. La principale raison pour laquelle le mécanisme de la déclaration de revenus a été rejeté était de fournir à l'intervieweur un instrument de collecte des données aussi simple que possible. On lui apprendrait plutôt à encourager le

répondant à consulter ses dossiers au besoin. Parallèlement, on continuerait à utiliser l'interviewer dépendante grâce aux indicateurs.

La première collecte des données de l'EDTR s'est déroulée en janvier (interview sur le travail) et en mai 1994 (interview sur le revenu). On s'est servi pour cela de l'IAO. Les taux de réponse pour l'interview de janvier sur le travail et l'interview de mai sur le revenu se sont respectivement établis à 86 % et à 76 %.

4.3 FAISABILITÉ DE LA MÉTHODE FISCALE

Vers le milieu de 1993, on a commencé à se demander si les données sur le revenu pourraient être recueillies par appariement avec les dossiers de Revenu Canada pour les répondants de l'enquête. Comme indiqué précédemment, pareille méthode allégerait considérablement le fardeau des répondants à l'égard des questions sur le revenu et améliorerait la qualité des données.

Dans le cadre des essais, on a posé à un sous-échantillon de répondants de l'Enquête sur la population active d'août 1993 une question visant à obtenir leur «permission». Une seule des deux parties du sous-échantillon avait déjà participé à l'EFC. La question devait établir si les répondants autoriseraient Statistique Canada à consulter leur dossier à Revenu Canada, au lieu de leur poser des questions sur leur revenu.

La question se lisait comme suit : «Nous voulons votre opinion concernant une nouvelle façon d'obtenir de l'information que Statistique Canada recueille présentement au moyen d'enquêtes. Nous cherchons une façon de réduire les coûts ainsi que le temps et l'effort que vous devez mettre pour répondre à nos enquêtes. Statistique Canada obtient actuellement les données sur le revenu en posant jusqu'à 25 questions sur les salaires, les pensions et d'autres genres de

revenu. La plupart des données figurent dans la déclaration d'impôt sur le revenu. Si vous deviez participer à une enquête de Statistique Canada sur le revenu, accepteriez-vous que l'on obtienne vos données directement de Revenu Canada?» Cette question a été posée à deux personnes de chaque ménage sélectionné. Les réponses par procuration n'ont pas été acceptées.

L'analyse portait sur des données non pondérées, car on s'intéressait plus au comportement des personnes échantillonnées qu'à celui de la population. L'échantillon se composait de 29 582 personnes, dont 17 % de non-répondants. Parmi les répondants, 59 % ont dit oui et 41 %, non. Les pourcentages étaient presque les mêmes pour les personnes qui avaient participé à l'EFC et pour les autres. Quand on examine les variables géographique ou démographiques, on note une variation légèrement plus importante, mais cette dernière n'est en aucun cas vraiment importante. La seule différence appréciable a été relevée chez les non-répondants de l'EFC, qui n'ont répondu oui que dans 42 % des cas.

Le taux de réponse à l'interview P et C de l'EFC dépassait les 80 %. Les résultats du test démontrent clairement qu'on assisterait à une forte baisse du taux de réponse si on recueillait exclusivement les données par appariement. Toutefois, la question posée n'était qu'hypothétique, et les intervieweurs n'avaient pas reçu pour instruction de déployer des efforts particuliers pour transformer un «non» en «oui». Les résultats auraient peut-être été différents si les personnes échantillonnées avaient vraiment dû choisir entre répondre à 25 questions ou permettre la consultation de leur dossier à Revenu Canada.

Les 42 % de non-répondants de l'EFC qui ont répondu «oui» suggèrent qu'on pourrait améliorer le taux de réponse en combinant l'interview à la méthode d'appariement.

La possibilité de recueillir les données sur le revenu par appariement avec la déclaration de revenus a été évaluée une fois de plus lors de la collecte des données sur le revenu de l'EDTR en mai 1994. On s'est pour cela servi d'une question similaire à celle de l'essai d'août 1993 -- «Vous avez peut-être remarqué que la déclaration d'impôt contient la plupart des informations que nous vous avons demandées durant cette interview. À votre avis, si [répondant] avait le choix l'an prochain, nous donnerait-il(elle) la permission d'obtenir cette information directement de Revenu Canada?» Dans ce cas, on a posé la question relative à l'autorisation immédiatement après le questionnaire sur le revenu de l'EDTR. À cette occasion, 56 % des répondants ont déclaré «oui», soit approximativement autant qu'au premier essai (59 %).

Dibbs et ses collaborateurs (1994) ont analysé les résultats des études de faisabilité. Il est également possible de se procurer un rapport sommaire de Dibbs, Poulin et Webber (1994) sur le même sujet.

4.4 APPROCHE MIXTE

Pour la deuxième vague de janvier et mai 1995, on a décidé de recourir à une approche mixte. Afin de parvenir aux résultats escomptés (qualité des données et allègement du fardeau des participants), il a été convenu qu'on proposerait aux répondants la consultation de leur déclaration de revenus. Les données sur le revenu des personnes accordant leur consentement seraient récupérées par une méthode d'appariement similaire à celle décrite à la section 4.2. Le numéro d'assurance sociale (NAS) - numéro de compte des particuliers qui présentent une déclaration de revenus - avait été obtenu pour faciliter l'extraction des données sur le revenu lors des vagues subséquentes. De cette façon, on veillerait à ce que les données longitudinales concernent la même personne, même si l'appariement peut

parfois être erroné. On n'a pas demandé directement aux répondants de donner leur NAS afin d'obtenir le consentement d'un maximum de personnes.

Par ailleurs, afin de maintenir un taux de réponse élevé, on a gardé l'approche de l'IAO avec carnet pour les répondants qui refusaient qu'on consulte leur déclaration de revenus.

La possibilité d'une consultation de la déclaration de revenus n'a pas été mentionnée lors de l'interview sur le travail de janvier. Avant l'entrevue du mois de mai, les répondants ont reçu un questionnaire - devant faciliter l'interview au téléphone - dans lequel on leur exposait la nouvelle option. Les personnes préférant cette approche devaient cocher une case, alors que les autres devaient remplir le questionnaire et le garder près du téléphone. Au moment de l'appel, l'intervieweur commençait par proposer la consultation de la déclaration de revenus. En cas de refus, l'interview se déroulait comme prévu. Afin de réduire l'incidence d'une couverture insuffisante par la déclaration de revenus, les répondants qui n'avaient pas présenté de déclaration pour l'année de référence ont subi l'interview. On a accordé grand soin à la rédaction des documents, et l'intervieweur a suivi un cours en vue d'accroître le plus possible les chances que le nombre de personnes consentantes dépasse celui relevé à l'essai initial.

Les résultats préliminaires du cycle de collecte des données de mai 1995 indiquent que 63 % des répondants ont accepté l'utilisation des données fiscales dans le cadre de l'EDTR. L'appariement sera irréalisable dans certains cas - il s'agira effectivement de non-répondants. Près de 4 % des personnes interrogées n'avaient pas produit de déclaration de revenus et 3 % n'ont pas répondu à la question concernant l'autorisation. Donc, 30 % ont refusé leur consentement. On a recueilli les données sur le revenu de ces trois derniers groupes (37 % des répondants) au moyen de l'enquête. (Les personnes qui n'avaient pas présenté de

déclaration de revenus ont aussi été interrogées afin de recueillir des données sur des petits revenus qu'elles ont pu recevoir.)

Lors des vagues subséquentes, les ménages où tous les répondants ont opté pour la consultation de la déclaration de revenus ne seront pas rappelés en mai.

5. IMPLICATIONS ET PROBLÈMES

De leur point de vue, les répondants pourraient se faciliter considérablement la tâche en acceptant qu'on consulte leur déclaration de revenus. En effet, le nombre total d'interviews du panel actuel passerait de 13 à 7. En outre, les interviews éliminées portent sur un sujet délicat et difficile à aborder, ce qui exige une certaine préparation. Le simple fait d'offrir un choix aux répondants laisse prévoir une hausse éventuelle du taux de réponse et un plus grand degré de satisfaction des répondants (et des intervieweurs).

Compte tenu des résultats des essais, on peut s'attendre à des données de qualité supérieure. Il est particulièrement important que les estimations d'un changement du revenu ou de la fréquence à laquelle se produisent de tels changements n'augmentent pas de façon artificielle à la suite des erreurs affectant les réponses. Grondin et Michaud (1994) ont constaté que pour un échantillon apparié, l'analyse des données sur la modification du revenu issues de l'enquête (répondants de l'EFC de 1992 et de l'essai de l'EDTR de mai 1993) et extraites de la déclaration de revenus produit des estimations nettement exagérées pour certaines variables, du côté de l'enquête. Il s'agit de la principale raison pour laquelle on a décidé d'utiliser le processus d'interview dépendante à l'aide d'indicateurs, préalablement testé en 1993, pour l'interview sur le revenu de 1994.

La mobilité économique des familles sera sans doute un des grands sujets de recherche auquel serviront les données sur le revenu de l'EDTR, surtout en ce qui concerne la stabilité du revenu et ses corrélats. Deux autres sujets comprennent la dynamique des faibles revenus et les liens entre le revenu familial et le comportement d'une personne sur le marché du travail. Des données sur le revenu des particuliers et des familles assorties d'un minimum d'erreurs affectant les réponses sont essentielles à de telles recherches.

On s'attend à ce que la consultation des déclarations de revenus réduise l'importance de la sous-déclaration pour certains types de revenu. Lors de l'interview, l'arrondissement du revenu peut soulever des difficultés que devrait aussi atténuer la consultation de la déclaration de revenus..

La couverture qu'offre la déclaration de revenus pose un autre problème. En 1994, près de 94 % des Canadiens de 20 ans et plus avaient rempli une déclaration de revenus. En proposant la consultation de cette dernière, on a demandé aux répondants s'ils en avaient remise une l'année précédente. Dans la négative, l'interview se poursuivait comme prévu. Cette méthode devrait réduire l'incidence d'une couverture insuffisante de la déclaration de revenus.

Lors de la première vague de mai 1994, le taux de réponse pour le revenu s'est chiffré à 77 %. Le recours à la méthode mixte devrait améliorer ce taux de réponse.

Avec une enquête longitudinale comme l'EDTR, l'érosion de l'échantillon demeure une préoccupation constante. Une non-réponse ou le refus de répondre à l'interview sur le revenu à une occasion pourrait inciter le répondant à en faire autant lors de l'interview sur le travail de janvier suivant. L'allègement du fardeau

des répondants par la consultation de la déclaration de revenus pourrait atténuer l'érosion de l'échantillon lors des vagues subséquentes.

Un autre grand problème concerne la fusion des données recueillies lors de l'interview avec celles tirées de la consultation des déclarations de revenus. Les deux sources présentent des différences théoriques et qualitatives. Mêler les deux sortes de données avec l'approche mixte comme on le fait dans l'EDTR est une nouveauté. On a toutefois trouvé une manière de résoudre la difficulté et on espère que cette méthode facilitera l'usage des données.

Tout d'abord, on prend les données sur le revenu telles qu'elles sont recueillies (avec quelques corrections) pour chaque répondant, qu'elles viennent de la consultation de la déclaration de revenus ou de l'enquête. On créera aussi des variables «fusionnées» sur le revenu, pour chaque répondant. Ces données seront un peu moins détaillées que dans chacune des deux sources. Dans la mesure du possible, la norme appliquée aux données fusionnées sera reprise lors de la collecte des données par interview. Les données fusionnées intégreront certains ajustements qui amélioreront la comparabilité des données des deux sources.

Les difficultés que posent les concepts mises à part, la section 2 soulignait que les données des régimes administratifs soulèvent un problème de stabilité. En effet, les règlements fiscaux évoluent avec le temps. Par conséquent, la définition des composantes du revenu peut changer. Ainsi, on peut réunir ou désunir certains éléments. On peut aussi ajouter ou supprimer des points à la déclaration de revenus. De tels changements peuvent prendre beaucoup d'importance avec le temps dans l'analyse des données longitudinales, qui gagne alors en complexité. Ces dernières années, la tendance a été d'ajouter de plus en plus de détails à la déclaration de revenus, y compris des sommes non imposables.

Nous n'avons pas encore abordé la question des coûts. En 1995, les possibilités d'économie étaient relativement faibles, principalement parce qu'il fallait encore entrer en contact avec tous les ménages pour l'interview sur le revenu. De plus, la consultation de la déclaration de revenus entraînera des frais de traitement supplémentaires. En 1995, près de 63 % des ménages ont autorisé l'appariement avec la déclaration de revenus. À compter de 1996, pour l'interview du mois de mai, on ne communiquera qu'avec les ménages qui s'y sont opposés, ce qui entraînera une économie de 160 000 \$ au niveau de la collecte des données.

La confidentialité est aussi un problème. La méthode de collecte mixte signifie qu'on remplacera directement les données d'enquête par celles de la déclaration de revenus et qu'on fusionnera ces dernières avec les autres données d'enquête. Puisque les données fiscales se retrouveront dans le fichier de microdonnées, les risques d'identifier une personne en seront augmentés, même si on arrondit les variables du revenu de l'EDTR.

6. EVALUATION

On a procédé à plusieurs évaluations en vue de résoudre divers problèmes. Ces évaluations sont reliées aux difficultés décrites à la partie précédente.

6.1 INCIDENCE POTENTIELLE SUR LE TAUX DE RÉPONSE

On a examiné les résultats de l'interview sur le revenu de mai 1994 pour déterminer quel effet une méthode de collecte mixte aurait sur le taux de réponse. Pour cela, il convient d'examiner deux grandes influences, une positive et l'autre négative. Tout d'abord, un certain pourcentage des non-répondants à l'interview sur le revenu pourraient accepter qu'on consulte leur dossier de Revenu Canada, ce qui augmenterait le taux de réponse. Ce résultat s'avérerait particulièrement

utile pour ceux qui ont répondu à l'interview sur le travail mais ont refusé de subir celle sur le revenu. D'un autre côté, peut-être ne pourra-t-on localiser la déclaration de revenus de certains répondants qui en ont autorisé la consultation (il pourrait s'agir de personnes qui n'ont pas présenté de déclaration, qui l'ont remise en retard ou pour lesquelles l'appariement n'a permis aucun lien); il s'ensuivrait donc une diminution du taux de réponse.

En 1994, 31 927 personnes étaient admissibles à l'interview sur le revenu. Sur ce nombre, 76 % (24 261) ont accepté de répondre aux questions. L'échantillon a été apparié aux déclarations de revenus de 1993 pour déterminer le taux de appariement. On a encore une fois procédé au appariement au moyen du nom, de la date de naissance, du sexe, de l'état matrimonial, du nom du conjoint et du code postal. Un lien avec la déclaration de revenus a pu être établi pour 85,1 % des répondants (20 637) et 76,2 % des non-répondants. (L'appariement a pu être réalisé pour les non-répondants grâce à la collecte des données démographiques pour la plupart d'entre eux lors d'une interview antérieure de l'EDTR). L'interview de mai 1994 comprenait la question hypothétique sur l'autorisation. Parmi les répondants, 56,7 % (11 702) des personnes appariées grâce à la déclaration de revenus et 53,3 % (1 932) de celles qui n'avaient pu l'être ont répondu «oui».

Avec la méthode de collecte mixte, on constate que l'incidence négative vient des répondants qui ont accordé leur permission, mais pour lesquels on n'a pu retrouver de déclaration de revenus. Le nombre de répondants diminue donc de 1 932 -- soit 6 % du taux de réponse. Nous savons maintenant que 727 de ces personnes n'avaient pas produit de déclaration de revenus pour l'année concernée. La baisse du taux de réponse attribuable à ces dernières aurait pu être évitée en demandant d'abord aux répondants s'ils avaient ou non produit une déclaration,

puis en recueillant les données sur le revenu des personnes qui répondent non, avec l'IAO.

Les tableaux 1 et 2 donnent la distribution des 1 932 répondants qui ont consenti à ce qu'on consulte leur déclaration de revenus, mais pour lesquels on ne l'a pas retrouvée, selon le revenu et le groupe d'âge. Le tableau 1 révèle qu'un nombre disproportionné de répondants avaient un faible revenu; par exemple, 74,8 % avaient un revenu inférieur à 10 000 \$. Le tableau 2 montre que les plus jeunes sont significativement sous-représentés. Il en va autant des plus âgés.

Tableau 1

Distribution du revenu (mentionné à l'interview de 1994 de l'EDTR) des répondants qui n'ont pu être appariés à la déclaration de revenus de 1993 mais qui en avaient autorisé la consultation.

Fourchette des revenus	Répondants(%)	Population(%)
selon l'enquête		
\$0- \$4999	55,4	11,8
\$5000- \$9999	19,4	13,4
\$10000- \$14999	8,5	15,2
\$15000- \$19999	4,5	11,4
\$20000- \$29999	4,6	17,9
\$30000- \$39999	3,6	12,5
\$40000- \$49999	1,6	7,8
\$50000 +	2,5	10,0

Tableau 2

Distribution selon l'âge des répondants qui n'ont pu être appariés à une déclaration de revenus en 1993 mais qui en avaient autorisé la consultation.

Groupe d'âge	Répondants(%)	Population(%)
16-19	28,9	6,9
20-24	6,6	9,0
25-34	10,1	21,6
35-44	7,5	21,1
45-54	12,4	15,6
55-64	13,0	11,2
65+	21,4	14,5

Au Canada, on se sert du numéro d'assurance sociale (NAS) comme numéro de compte pour la déclaration de revenus des particuliers. Il s'ensuit que l'absence de NAS prédit un faible revenu --- assez faible pour ne pas être imposable. Parmi les 1 932 répondants pour lesquels on a pas retrouvé de déclaration de revenus en 1993, l'EDTR détenait un NAS pour 727 d'entre eux tandis que pour les 1 205 autres elle n'en avait pas.

Dans le groupe sans NAS, 62 % des répondants étaient célibataires, 54 % étaient des femmes et 45 % avaient moins de 20 ans. Dans l'ensemble, 68 % avaient gagné moins de 10 000 \$. Parmi ceux de moins de 20 ans, 98 % signalaient un revenu inférieur à 10 000 \$.

Les personnes pour lesquelles on possédait le NAS constituent un groupe très différent. Ainsi, les femmes mariées en représentent 79 %. Ces dernières sont plus âgées, puisque 58 % avaient au-delà de 54 ans. Un plus grand nombre d'entre elles ont aussi un revenu inférieur à 10 000 \$ (86 %).

Manifestement, beaucoup de personnes susceptibles de devenir un non-répondant avec la méthode de collecte mixte gagnent un faible revenu, qu'elles aient ou non un NAS. Bon nombre de personnes pour lesquelles l'EDTR n'a pas de NAS sont des jeunes et/ou des célibataires qui n'ont pas encore une grande expérience du marché du travail. À l'opposé, celles détenant un NAS ont tendance à être plus âgées. Il s'agit dans la majorité des cas de femmes mariées. Pour le cycle de 1995 de l'EDTR, notons que l'imputation sera plus facile si jamais ces personnes se désistent puisqu'elles ont accepté de répondre à l'enquête de 1994.

Les personnes qui refusent de répondre à l'entrevue sur le revenu mais consentent à ce qu'on consulte leur déclaration de revenus auront une influence positive sur le taux de réponse, ce qui pourrait contrebalancer la situation. Si on se fie aux résultats de l'essai d'août 1993, environ 42 % des personnes qui n'avaient pas voulu répondre à l'enquête sur le revenu ont déclaré être disposées à autoriser la consultation de leur déclaration de revenus. Si on applique ce taux au nombre de non-répondants qu'on a pu apparier avec une déclaration, le taux de réponse de l'EDTR pourrait augmenter de 7 %.

Bien que les deux effets semblent approximativement s'annuler, le résultat net sur le taux de réponse n'est pas clair (du moins pour les interviews complètes). L'existence de nombreuses personnes pour lesquelles on n'a pu trouver de déclaration de revenus, même si elles en permettent la consultation, ne paraît pas poser de problème majeur. On appréciera mieux la difficulté lors de l'enquête de 1995, car on demandera alors de façon explicite si la personne a présenté une déclaration de revenus ou non. Dans le second cas, les données sur le revenu seront recueillies par interview.

6.2 QUALITÉ DES DONNÉES DÉTAILLÉES

Pour évaluer la qualité des données, on a apparié l'échantillon de l'EDTR de 1994 et les déclarations de revenus, comme on l'a fait lors de l'essai de 1993. On a ainsi pu procéder à une série de comparaisons articulées sur trois grandes dimensions :

1) la comparaison des répondants et des non-répondants, afin d'évaluer l'impact qu'ont les non-répondants qui autorisent la consultation de leur déclaration de revenus (en d'autres termes il faut se poser les questions suivantes : existe-t-il des différences entre les répondants et les non-répondants et pourra-t-on améliorer la qualité des données en ajoutant quelques non-répondants à l'échantillon par la consultation de leur déclaration de revenus?);

2) la comparaison des données fiscales et des données d'enquête des «bons» répondants, en vue de préciser la variation des concepts entre l'enquête et la déclaration de revenus. Par «bons» répondants, on entend ceux qui n'ont pas répondu qu'ils «ne savaient pas» et qui n'ont pas refusé de répondre aux questions sur le revenu, dont les données ont franchi toutes les règles de validation et pour lesquels on a retrouvé une déclaration de revenus;

3) la comparaison des données fiscales et des données d'enquête pour l'ensemble de l'échantillon, après traitement, correction et imputation, afin d'évaluer l'effet global d'une méthode mixte de collecte des données sur les utilisateurs.

Nous ne nous intéresserons ici qu'aux deux premières dimensions.

Les comparaisons se limitent à six types de revenu : salaires et traitements (ST), revenu d'un travail autonome agricole et non agricole (TA, TNA), intérêts et dividendes (I), assurance-chômage (AC) et aide sociale (AS).

On a choisi les salaires et traitements, le revenu d'un travail autonome agricole et le revenu d'un travail autonome non agricole parce que la manière dont le revenu d'un travail autonome doit être déclaré diffère selon le genre de travail (à savoir, travailleur incorporé ou non-incorporé).

On a également retenu les intérêts et les dividendes car ils ont tendance à être sous-déclarés dans les enquêtes ayant pour but de recueillir des données sur le revenu.

Les prestations d'assurance-chômage ont été sélectionnées parce qu'elles aussi ont tendance à être sous-déclarées dans une certaine mesure au moment de l'enquête. En ce qui concerne l'EDTR cependant, on demande à deux reprises si la personne touche des prestations : une fois lors de l'entrevue sur le travail, où le répondant est prié de dire s'il a reçu des prestations d'assurance-chômage en général et, dans l'affirmative, quand il a reçu ces prestations ainsi qu'une autre fois lors de l'enquête sur le revenu, où il doit indiquer le montant des prestations touchées. On espère qu'en sollicitant cette information à deux reprises et en utilisant l'interview dépendante, on pourra améliorer la réponse à cette question lors de l'interview sur le revenu.

Enfin, on s'est penché sur l'aide sociale, car cette dernière soulève un problème particulier. En effet, la question est posée au niveau de la personne. Cependant, les paiements d'AS sont calculés d'après la composition du ménage et, quoi qu'ils ne soient versés qu'à une seule personne, ils aident tout le ménage.

6.2.1 Comparaison des répondants et des non-répondants à l'IAO

Le tableau 3 compare les données fiscales des répondants et des non-répondants qu'on a pu appairier à la déclaration de revenus pour les six sources de revenu précitées.

À la section 6.1, on a montré que beaucoup de personnes dont on n'a pas retrouvé la déclaration de revenus avaient un revenu assez faible, si bien qu'elles pourraient différer quelque peu des personnes pour qui on a pu établir un appariement. Faute de données fiscales (qui n'existaient peut-être pas), ces personnes n'apparaissent pas au tableau 3. Le tableau ne donne donc qu'une comparaison partielle des répondants et des non-répondants. Quoiqu'il en soit, il est possible de tirer certaines conclusions générales.

La proportion de répondants et de non-répondants qui signalent chaque source de revenu diffère habituellement peu. Un pourcentage plus élevé de non-répondants semble néanmoins avoir reçu de l'aide sociale ou un revenu d'un travail autonome (surtout non agricole) .

Bien que les sources de revenu sur la déclaration de revenus des répondants et des non-répondants soient similaires, le tableau 4 révèle que la moyenne et la médiane des montants sont généralement plus élevées pour les non-répondants.

Les deux tableaux indiquent une légère différence entre les répondants et les non-répondants pour ce qui est des sources de revenu signalées. En outre, lorsqu'un montant est déclaré, il a tendance à être plus élevé, parfois de beaucoup, chez les non-répondants. Une partie du biais pourra être corrigée si certains non-répondants autorisent la consultation de leur déclaration de revenus.

Tableau 3

Sources de revenu (selon la déclaration de revenus) rapportées par les répondants et les non-répondants pour lesquels l'appariement a pu être établi.

	ST		TA		TNA	
	%	n	%	n	%	n
Rép.	67,9	14021	3,6	737	8,4	1734
NR	67,8	3960	4,0	231	9,6	561

	I		AC		AS	
	%	n	%	n	%	n
Rép.	39,5	8143	19,6	4053	7,1	1472
NR	39,6	2311	20,2	1177	9,1	533

Tableau 4

Revenu moyen et revenu médian rapportés (selon la déclaration de revenus), par source de revenu et type de répondants.

		ST	TA	TNA	I	AC	AS
Moyen	Rép.	\$24061	\$4532	\$10781	\$2497	\$5293	\$5845
	NR	\$24175	\$6143	\$13282	\$3388	\$5277	\$6103
Médian	Rép.	\$19724	\$1249	\$3690	\$467	\$4364	\$5662
	NR	\$18142	\$2726	\$4495	\$634	\$4344	\$5382

6.2.2 Comparaison des données d'enquête et des données fiscales des «bons» répondants

On a procédé à des études plus poussées et comparé les données d'enquête des «bons» répondants à leurs données fiscales. Le but de la comparaison était de déterminer dans quelle mesure la définition des sources de revenu était comparable

entre l'EDTR et la déclaration de revenus. Pour cela, on a dû redéfinir certaines catégories pour les rendre «comparables».

Certains revenus non imposables, comme l'aide sociale, apparaissent sur la déclaration de revenus, mais pas d'autres. Il suffit de penser à la pension d'ancien combattant, aux héritages et aux prix de loterie. Pour ces sommes, la déclaration de revenus n'est d'aucune utilité. Il faut poser des questions spécifiques. Parmi les 15 862 «bons» répondants de 1994, seulement 340 ont signalé des sommes non imposables qui n'apparaîtraient pas sur la déclaration de revenus. Dans l'ensemble, le montant moyen mentionné par les «bons» répondants variait très peu, soit environ 0,3 % du revenu global. Le revenu moyen et le revenu médian des 340 qui ont rapporté un tel montant représentaient respectivement 18,0 % et 9,5 % du revenu total. Seulement environ 9 % des 340 répondants tiraient 50 % ou plus de leur revenu de la source en question.

En règle générale, les deux sources de données présentent des différences conceptuelles. Le but de l'exercice était de déterminer dans quelle mesure certaines de ces différences sont signalées dans la réalité. Il pourrait s'agir notamment du revenu d'un «travail au noir» ou de petites sommes issues de revenus d'intérêt (on n'émet pas de reçus d'impôt pour les sommes inférieures à 100 \$), qu'une personne pourrait théoriquement rapporter dans le cadre d'une enquête, mais ne pas inscrire sur sa déclaration.

Le tableau 5 montre qu'il existe un fort écart dans la déclaration des intérêts et des dividendes, même chez les «bons» répondants. On rapporte aussi un revenu plus important dans le cadre d'un travail autonome sur la déclaration de revenus.

Les moyennes et les médianes qui apparaissent au tableau 6 sont plus élevées pour les données de l'enquête, surtout pour le revenu d'un travail autonome et les

intérêts. On peut en conclure que les montants non signalés lors de l'enquête sont habituellement faibles.

Puisqu'on a pu trouver la déclaration de revenus de tous les «bons» répondants, on a décidé de procéder à une micro-comparaison du taux de «concordance» des données entre les deux sources. Dans ce contexte, concordance indique seulement s'il y a eu déclaration ou non du montant dans chaque cas. Le tableau 7 présente le résultat de la comparaison. Les quatre premières lignes révèlent la fréquence à laquelle la source de revenu n'a pas été déclarée dans l'un et l'autre cas (=0 les deux), a été signalée dans les deux cas (>0 les deux), ou n'a été rapportée que dans la déclaration de revenus (>0 impôt) ou lors de l'enquête (>0 enquête). La dernière ligne donne la proportion de répondants qui ont mentionné un montant dans les deux cas après avoir signalé un revenu à l'une ou l'autre occasion, ou aux deux.

Tableau 5

Sources de revenu rapportées dans la déclaration de revenus et l'EDTR (1994) par les «bons» répondants

	ST		TA		TNA	
	%	n	%	n	%	n
Impôt	65,2	10338	2,8	440	7,3	1158
Enquête	65,2	10336	2,3	360	5,1	805

	I		AC		AS	
	%	n	%	n	%	n
Impôt	38,4	6084	16,5	2614	6,6	1043
Enquête	29,7	4391	15,4	2449	6,7	1066

Tableau 6

Revenu moyen et revenu médian des «bons» répondants ayant déclaré un montant, selon le type de revenu et la source des données.

		ST	TA	TNA	I	AC	AS
Moyenne	Impôt	\$25092	\$6446	\$12339	\$2542	\$5386	\$6091
	Enquête	\$25060	\$9763	\$14635	\$2912	\$5197	\$5882
Médiane	Impôt	\$21261	\$2292	\$3522	\$487	\$4505	\$5905
	Enquête	\$21353	\$5202	\$5482	\$597	\$4247	\$5710

Tableau 7

Micro-comparaison des données d'enquête et des données fiscales des «bons» répondants.

	ST	TA(%)	TNA(%)	I(%)	AC(%)	AS(%)
=0 les deux	31,1	96,1	90,3	57,5	82,5	91,8
>0 les deux	64,1	2,5	4,3	27,9	15,5	5,7
>0 impôt	2,4	1,0	3,8	12,6	1,5	1,2
>0 enquête	2,4	0,4	1,6	2,0	0,5	1,3
% concordance	92,9	63,8	44,2	65,6	88,8	69,5
(>0)						

Plusieurs conclusions intéressantes ressortent du tableau 7. Il se pourrait qu'une partie de l'économie clandestine soit signalée lors de l'enquête, mais pas dans la déclaration de revenus. Pour la plupart des catégories, un montant apparaît sur la déclaration de revenus mais n'est pas mentionné au moment de l'enquête. Il existe un très faible taux de concordance entre les données de l'enquête et les données fiscales pour le revenu tiré d'un travail autonome, et il convient d'approfondir les recherches pour déterminer si les sommes mentionnées apparaissent ailleurs dans le questionnaire. Les intérêts et les dividendes sont aussi sous-déclarés lors de l'enquête. Bien qu'on note un meilleur taux de concordance pour les prestations d'assurance-chômage, il est intéressant de remarquer que, même en posant une question à ce sujet deux fois dans le cadre de l'enquête et en utilisant l'interview

dépendante pour remédier au problème d'une mémoire défaillante, on obtient une sous-estimation de 8,8 % sous réserve que les prestations mentionnées sur la déclaration de revenus représentent la «vérité». Enfin, le taux de concordance est plutôt faible pour l'aide sociale, et il semble exister un problème de disparité entre les résultats de l'enquête et la déclaration de revenus. La situation pourrait être attribuable au fait que l'aide sociale n'est accordée qu'à un seul membre du ménage, tel qu'indiqué précédemment, bien que plusieurs personnes y soient admissibles.

7. CONCLUSIONS

La méthode mixte de collecte des données sur le revenu combinant les dossiers administratifs de l'impôt aux données de l'enquête paraît très prometteuse -- assez pour qu'on ait décidé d'y recourir afin de recueillir les données sur le revenu de 1994, lors de l'entrevue de mai 1995 de l'EDTR. On estime que l'approche combinée allégerait non seulement le fardeau des répondants, mais améliorerait la qualité des données de l'EDTR. Néanmoins, il convient d'approfondir les recherches pour évaluer exactement l'impact de la nouvelle méthode.

Il est impossible de retrouver la déclaration de revenus d'un certain nombre de personnes qui acceptent qu'on consulte leur dossier. Dans la plupart des cas, ces personnes paraissent avoir un faible revenu. Dans quelques autres cas, il est impossible de repérer le dossier, même avec le numéro d'assurance sociale. Il s'ensuit que la personne n'a pas produit de déclaration de revenus. Puisque l'EDTR est une enquête longitudinale, on pourrait toujours ajouter les données fiscales si les personnes en question produisaient une déclaration de revenus subséquemment.

Les non-répondants ne diffèrent pas considérablement des répondants en ce qui concerne les montants déclarés pour diverses sortes de revenu; les montants moyen et médian rapportés par les non-répondants sont habituellement plus élevés que ceux des répondants. La qualité des données sur le revenu de l'EDTR devrait donc s'améliorer si on relève le taux de réponse en amenant certains non-répondants à accepter qu'on consulte leur dossier fiscal, ce qui en fera des répondants.

La comparabilité des données fiscales et des données d'enquête est passablement bonne au niveau macroscopique. Toutefois, la micro-analyse ne donne pas un taux de concordance particulièrement élevé, même quand la comparaison se restreint à ce qu'on appelle les «bons répondants». Dans la plupart des cas de disparité, un montant apparaît sur la déclaration de revenus mais n'est pas mentionné à l'enquête. Il se pourrait que certains aspects de la déclaration de revenus ne se retrouvent pas dans l'enquête (ou soient mentionnés ailleurs), ou que l'enquête suscite une sous-déclaration.

Certaines variables comme l'aide sociale devront peut-être être dérivées au niveau du ménage, à moins qu'on doive effectuer un nouveau traitement au niveau de l'individu (en calculant la moyenne des montants) pour une plus grande uniformité, surtout dans le cas de l'enquête longitudinale.

Notons qu'en 1995, la question sur l'autorisation a été posée en mai. En acceptant, le répondant pouvait abrégé l'entrevue sans toutefois pouvoir y échapper. Lors du cycle de 1996, les 37 % de répondants qui n'avaient pas accordé leur permission en 1995 devront répondre à la même question. Cette fois cependant, elle sera posée en janvier, ce qui donnera l'opportunité aux répondants d'éviter l'interview de mai en entier. Pareillement, la documentation envoyée en

janvier signalera aux personnes qui ont déjà accepté qu'elles peuvent revenir sur leur décision

REMERCIEMENTS

Nous remercions sincèrement Ruth Dibbs, Elaine Fournier et Maryanne Webber pour avoir défini les sources de revenu qui nous ont permis d'effectuer les comparaisons. Nous tenons aussi à remercier Philip Giles pour ses commentaires utiles sur une première ébauche du document.

BIBLIOGRAPHIE

Brackstone, G. (1987). Utilisation des dossiers administratifs à des fins statistiques. *Techniques d'enquête*, 13, 35-51.

Dibbs, R., Fournier, E., Grondin, C., Leesti, T., Poulin, S., Saint-Pierre, Y., Webber, M. (1994). Utilisation des données fiscales dans l'Enquête sur la dynamique du travail et du revenu: Étude de faisabilité. Statistique Canada, document de recherche de l'EDTR 94-11.

Dibbs, R., Poulin, S., Webber M. (1994). Utilisation des données fiscales dans l'Enquête sur la dynamique du travail et du revenu : Rapport sommaire. Statistique Canada, document de recherche de l'EDTR 94-11.

Giles, P. (1993). Le questionnaire et les procédures de collecte des données sur le revenu de l'EDTR - mai 1993. Statistique Canada, document de recherche de l'EDTR 93-04.

Grondin, C., Michaud, S. (1994). Qualité des données sur le revenu recueillies selon la méthode de l'interview assistée par ordinateur pour l'EDTR. Statistique Canada, document de recherche de l'EDTR 94-15.

Poulin, S. (1993). L'utilisation des données de l'impôt sur le revenu pour l'EDTR. Statistique Canada, document de recherche de l'EDTR 93-01.