

N° 98-05 au catalogue

**ASPECTS GÉNÉRAUX DE L'ENQUÊTE SUR LA
DYNAMIQUE DU TRAVAIL ET DU REVENU**

Numéro d'enregistrement du produit 75F0002M

Mars 1998

Mylène Lavigne, Division des méthodes d'enquêtes sociales

Sylvie Michaud, Division des méthodes d'enquêtes sociales

La série des documents de travail sur la Dynamique du revenu et du travail est conçue en vue de communiquer les résultats des études ainsi que les décisions importantes ayant trait au Programme sur la dynamique du revenu et du travail. Elle est une continuation de la Série de documents de recherche de l'EDTR. Ces documents sont disponibles en français et en anglais. Pour obtenir une description sommaire des documents disponibles ou un exemplaire de ces documents, communiquez avec l'Unité de diffusion, Édifice Jean-Talon, 7^e étage, section B5, Statistique Canada, Ottawa (Ontario), Canada, K1A 0T6; par INTERNET: DYNAMIQUE@STATCAN.CA; par téléphone au (613) 951-7355 ou sans frais au 1-888-297-7355; ou par télécopieur au (613) 951-3012.

SOMMAIRE

L'enquête sur la dynamique du travail et du revenu (EDTR) est une enquête de Statistique Canada qui a pour objet de servir à la recherche sur les changements qui interviennent dans le temps quant à la situation des Canadiens vis-à-vis de l'activité et à leur bien-être économique. Deux caractéristiques importantes du plan de l'enquête découlent directement de cet objectif. Premièrement, l'EDTR est une enquête longitudinale ; chaque panel participe à l'enquête pendant six ans. Deuxièmement, l'EDTR vise des ménages entiers et porte sur un éventail de sujets suffisamment large pour que l'on puisse recueillir des données sur des situations familiales et des événements démographiques importants. Cet aspect de l'enquête permettra d'examiner les liens qui existent entre des événements démographiques, les comportements d'activité et le revenu. L'article donne un aperçu des principaux buts de l'enquête et de la méthodologie employée.

TABLE DES MATIÈRES

	Page
1. Introduction	1
2. Méthodologie de l'EDTR	1
2.1 Sélection et caractéristiques de l'échantillon	1
2.2 Plan de renouvellement de l'échantillon	2
2.3 Collecte des données	4
2.4 Règles de suivi	6
2.5 Dépistage avec l'IAO	14
2.6 Taux de réponse de l'EDTR	16
2.7 Imputation	18
2.8 Pondération	20
2.9 Confidentialité	24
3. Conclusion	24
Bibliographie	26

1. Introduction

L'enquête sur la dynamique du travail et du revenu (EDTR) est une enquête longitudinale. Elle ajoute une nouvelle dimension aux données d'enquêtes existantes en matière d'activité sur le marché du travail et de revenu : les changements qui touchent les personnes au cours d'une période donnée. Au coeur des objectifs visés par l'enquête, on trouve le désir de comprendre le bien-être économique des Canadiens : à travers quels changements économiques doivent passer les personnes et les familles, et quel rôle jouent à cet égard les changements touchant le travail rémunéré, la composition de la famille, la réception de paiements de transfert gouvernementaux, ou d'autres facteurs. Par son aspect longitudinal, l'enquête permet de percevoir ces faits concomitants et souvent reliés entre eux. Contrairement à certaines enquêtes longitudinales, l'EDTR recueille des données sur tous les membres des ménages afin de bien cerner les événements familiaux importants qui surviennent au cours des six ans. L'EDTR est donc la première enquête-ménage menée par Statistique Canada à fournir des données nationales sur la stabilité du revenu d'une famille ou d'une personne type. Elle permettra ainsi d'avoir une meilleure idée de la nature et de l'étendue de la pauvreté au Canada.

2. Méthodologie de l'EDTR

2.1 Sélection et caractéristiques de l'échantillon

L'EDTR est une enquête-ménage. L'échantillon de personnes d'un panel est choisi parmi l'échantillon de l'Enquête sur la Population Active (EPA) [1]. L'échantillon de l'EPA est prélevé d'une base aréolaire selon un plan de sondage probabiliste à plusieurs degrés. L'unité d'échantillonnage au dernier degré est le logement.

L'EPA vise la population des 10 provinces, à l'exception des résidents des réserves indiennes, des membres à temps plein des Forces armées canadiennes et des pensionnaires d'établissements institutionnels. La population visée par l'échantillon initial de l'EDTR est la même, à l'exception près des membres des Forces armées qui vivent à l'extérieur des casernements.

Quand un panel est introduit dans l'échantillon, toutes les personnes faisant partie des ménages choisis au moment de la sélection du panel, indépendamment de l'âge de ces personnes, deviennent des membres de l'échantillon longitudinal de l'EDTR. Ces personnes demeurent des membres de l'échantillon longitudinal pendant toute la période de participation du panel à l'enquête, même si elles déménagent.

Aucune autre personne ne devient *membre de l'échantillon longitudinal+ pendant la période de participation de ce panel à l'enquête. Ainsi, pour chaque panel, l'échantillon longitudinal est constitué au moment de l'introduction du panel (c'est-à-dire en janvier de la première année de référence visée dans le cas de ce panel) et reste tel quel pendant toute la durée du panel.

2.2 Plan de renouvellement de l'échantillon

Bien que des estimations annuelles (transversales) soient aussi produites, l'objectif principal de l'enquête est évidemment l'analyse longitudinale. Pour l'analyse longitudinale, la durée des périodes pendant laquelle les unités font partie de l'échantillon doit être aussi longue que possible. Toutefois, la fiabilité des estimations transversales diminuerait chaque année en raison des variations de la population et de la perte d'effectifs de l'échantillon. Il est donc nécessaire de procéder à un certain renouvellement de l'échantillon. On a étudié plusieurs plans de renouvellement de l'échantillon et on a retenu celui-ci : le premier panel a

commencé à participer à l'enquête en janvier 1993 et le deuxième sera introduit dans l'échantillon en janvier 1996 (voir le tableau 1).

Ces deux panels participeront à l'enquête de 1993 à 1998 et de 1996 à 2001, respectivement. Une fois le deuxième panel introduit dans l'échantillon, celui-ci aura atteint sa taille prévue et définitive, soit environ 30 000 ménages. En janvier 1999, un autre panel sera introduit (pour les années de référence 1999 à 2004), afin de remplacer le premier. Selon ce plan, un panel sera remplacé par un autre à tous les trois ans. La taille du premier panel est de 15 000 ménages. Ce nombre représente environ 31 000 personnes âgées de 16 ans et plus qui ont à répondre à des questions sur leur travail et leur revenu de l'année précédente. La taille des prochains panels sera d'environ 15 000 ménages.

TABLEAU 1

PLAN DE RENOUVELLEMENT DE L'ÉCHANTILLON														
Panel	Année													
	93	94	95	96	97	98	99	00	01	02	03	04	05	06
1	P	I	I	I	I	I	I							
2				P	I	I	I	I	I	I				
3							P	I	I	I	I	I	I	
4										P	I	I	I	I
<i>P = Interviews préliminaires I = Interviews sur le travail et le revenu</i>														
<i>Chaque panel contient au départ environ 15 000 ménages</i>														

2.3 Collecte des données

Quand un panel est introduit dans l'échantillon, une interview préliminaire est réalisée en janvier auprès de chaque membre âgé de 15 ans et plus des ménages choisis. Les six années subséquentes, en commençant par l'année suivant celle où le panel est introduit dans l'échantillon, on recueille des données sur le travail et le revenu. Il est à noter qu'on communiquera avec un *répondant longitudinal+ treize fois en sept ans : c'est-à-dire la première année pour l'interviewer préliminaire et chacune des six années suivantes pour les collectes des données sur le travail et le revenu. La collecte des données sur le travail a lieu en janvier, tandis que celle des données sur le revenu est effectuée en mai. Les données sur le travail sont recueillies aussitôt que possible après la fin de l'année de référence de façon à réduire l'impact des erreurs de mémoire. La raison principale pour laquelle les données sur le revenu sont recueillies en mai est que, de cette façon, les répondants peuvent consulter leurs déclarations qu'ils viennent de produire pour l'impôt sur le revenu. Toutefois, depuis la collecte de mai 1995, on propose aux répondants le choix de répondre à l'interview sur le revenu en mai ou de donner la permission à Statistique Canada d'utiliser leur déclaration d'impôt. Pour les gens qui nous donneront la permission d'utiliser leur déclaration d'impôt, on fera un appariement statistique pour aller chercher leurs données sur le fichier d'impôt et ils ne seront plus interviewés en mai. Les variables nom, prénom, date de naissance, sexe, état matrimonial, adresse, code postal et prénom du conjoint sont utilisées pour effectuer l'appariement statistique.

Pour les collectes des données sur le travail et le revenu, seules seront interviewées les personnes âgées de 16 ans et plus faisant partie de l'échantillon longitudinal et les *cohabitants+ des ménages (c'est-à-dire les personnes qui habitent avec un membre de l'échantillon longitudinal au 1^{er} janvier de l'année de l'interview).

Les données de l'EDTR sont recueillies selon la méthode de l'interview assistée par ordinateur (IAO). La collecte des données est décentralisée : les intervieweurs travaillent généralement à leur domicile, mènent les interviews par téléphone et transmettent les données par modem au bureau régional de Statistique Canada le plus proche. (Par mesure de sécurité, les résultats de l'enquête sont cryptés lors de la transmission.) La déclaration par personne interposée est acceptée dans la mesure où le répondant substitut connaît suffisamment l'enquêté pour répondre aux questions. Autrement, l'intervieweur prend des dispositions pour rappeler l'enquêté.

L'EDTR a été conçue en fonction de la méthode de l'IAO et exploite les possibilités de celle-ci pour améliorer la qualité des données. Par exemple, il faut obtenir de nombreuses dates dans le cadre de l'interview sur le travail telles : les dates des périodes d'emploi, les dates des périodes sans emploi, les dates des interruptions de travail, etc. L'IAO permet de vérifier la cohérence des renseignements relatifs à ces dates au moment où ils sont fournis par le répondant et de demander des précisions dans les cas où il semble y avoir des lacunes ou des incohérences. Cette vérification interactive des dates se traduit par des données de bien meilleure qualité.

L'IAO facilite de plus le recours à l'interview avec rétro-information, technique qui consiste à rappeler au répondant des renseignements qu'il a fournis lors de l'interview précédente, pour l'aider à se souvenir des événements qui sont survenus depuis. Une enquête comme l'EDTR peut donner lieu à des *problèmes de lisière+, parce que les répondants ont parfois de la difficulté à se souvenir des dates de leur entrée en fonction, de leur cessation d'emploi, de leurs interruptions de travail, etc. sur une période d'un an. Ces erreurs de mémoire peuvent entraîner une proportion exagérément élevée de périodes commençant ou se terminant à la *lisière+ de deux

périodes de référence consécutives. Le fait de rappeler au répondant certains renseignements recueillis lors de l'interview précédente peut aider à atténuer ces problèmes de lisière. Cette technique est beaucoup plus facile à utiliser dans un environnement IAO. L'IAO permet en outre de déceler les erreurs de cohérence logique entre les renseignements recueillis dans l'interview sur le travail et, quelques mois plus tard, ceux recueillis dans l'interview sur le revenu.

L'intervieweur demande alors des précisions.

Enfin, la méthode de l'IAO permet de tenir une liste des personnes avec lesquelles le répondant a vécu depuis le début du panel. Si, quelques années après le début du panel, le répondant emménage avec un soi-disant nouveau cohabitant, l'intervieweur vérifie la liste de tous les anciens membres du ménage pour voir si ce cohabitant ne serait pas en fait un membre du ménage de retour après une absence. (Distinguer un membre de retour d'un vrai nouveau membre du ménage peut parfois être étonnamment complexe bien que cela ne semble être qu'une simple formalité.) On utilise à peu près la même approche dans l'interview sur le travail, afin de distinguer les véritables nouveaux employeurs des employeurs pour lesquels le répondant retourne travailler après ce qui semblait une cessation d'emploi permanente.

2.4 Règles de suivi

La réalisation d'enquêtes longitudinales exige l'établissement de règles pour déterminer quelles personnes doivent faire l'objet d'un dépistage et lesquelles doivent être interviewées tout au long de l'enquête. Selon la conception, les objectifs et le budget de l'enquête, ces règles peuvent être très simples ou très complexes. Dans le cas de l'EDTR, les règles de suivi sont relativement simples en principe, mais plutôt complexes d'un point de vue opérationnel. La complexité des

règles relatives aux personnes qui doivent faire l'objet d'un dépistage et à celles qui doivent être interviewées augmente graduellement dans les premiers cycles d'interviews. On les expose ci-après en ordre chronologique.

Premier cycle d'interviews sur le travail

Entre janvier 1993 et janvier 1994, certains répondants longitudinaux auront déménagé et certains cohabitants se seront joints à l'échantillon. Les cas à considérer sont les suivants :

- Les répondants longitudinaux âgés de 16 ans et plus qui ont déménagé dans un autre logement du champ de l'enquête¹ sont dépistés à leur nouvelle adresse et interviewés pour la collecte des données sur le travail.
- Les répondants longitudinaux âgés de 16 ans et plus qui ont déménagé dans les Territoires, dans un camp militaire, dans une réserve indienne ou aux États-unis (endroits qui sont tous exclus du champ de l'enquête) sont dépistés et interviewés pour la collecte des données sur le travail comme les répondants qui déménagent dans un autre logement du champ de l'enquête.
- Dans le cas des répondants longitudinaux âgés de 16 ans et plus qui sont partis vivre dans un établissement institutionnel (pour plus de six mois) ou à l'étranger, on consigne leur nouvelle adresse. Lors des prochaines collectes, on vérifiera leur adresse courante auprès

¹ Le terme *champ* s'applique à un logement qui faisait partie de la base originale de l'EPA.

de la personne ressource ou d'un autre membre du ménage en vue de reprendre les interviews à leur retour. (Il est à noter que l'intervieweur doit déterminer si une personne qui est partie vivre dans un établissement de soins réside réellement dans un *établissement institutionnel+ mais que d'autres considèrent comme des maisons de chambres ou des maisons de pension spécialisées. La règle est la suivante : dans le doute, il faut dépister et interviewer.)

- Dans le cas d'un répondant longitudinal décédé, on ne recueille d'autres renseignements que la date de son décès. D'un point de vue opérationnel, la personne est éliminée de l'échantillon, mais les renseignements qui la concernent sont conservés dans le fichier de données.

- Certains répondants longitudinaux résideront désormais avec des cohabitants. Une personne qui a emménagé chez un répondant longitudinal après janvier 1993 est un *cohabitant+ et fait partie de l'échantillon de l'EDTR tant et aussi longtemps qu'elle habite avec un répondant longitudinal. Lors du premier cycle d'interviews sur le travail tous les cohabitants sont forcément des *nouveaux membres+, c'est-à-dire des personnes qui sont interviewées pour la première fois dans le cadre de l'EDTR. On réalise auprès de tout nouveau membre âgé de 16 ans et plus une version abrégée de l'interview préliminaire (l'interview préliminaire abrégée), en plus de l'interview pour la collecte des données sur le travail.

- À l'occasion, il est possible qu'un répondant longitudinal ait déménagé dans un logement collectif non institutionnel. La règle générale selon laquelle il faut interviewer tous les cohabitants peut alors poser un problème, par exemple, la personne a déménagé dans un monastère où résident cinquante autres moines. Pour éviter ce problème, seuls les membres de la famille économique (c'est-à-dire toute personne qui lui est apparentée par le sang, par alliance, par union libre ou par adoption) d'une personne ayant déménagé dans un logement collectif non institutionnel seront considérés comme des cohabitants et inclus dans l'EDTR.
- Toutes les personnes âgées de 15 ans, qu'il s'agisse de répondants longitudinaux ou de cohabitants, doivent répondre à une interview préliminaire en vue de leur première participation à la collecte des données sur le travail en 1995.
- Les répondants longitudinaux âgés de moins de 15 ans qui déménagent sont dépistés suivant les mêmes règles que celles qui s'appliquent dans le cas des personnes plus âgées. Les renseignements recueillis à leur sujet, et ceux qui concernent les jeunes cohabitants, se limitent aux caractéristiques démographiques de base et à l'année d'études.

Premier cycle d'interviews sur le revenu

En mai 1994, on a recueilli auprès du premier panel la première série de données sur le revenu. C'est pour permettre aux répondants de consulter leurs documents fiscaux que l'on recueille les renseignements en mai. En fait, cette collecte n'est

réalisée séparément de celle de janvier que pour des raisons pratiques ; il ne s'agit pas vraiment d'un cycle distinct. Comme les données sur le travail, celles qui portent sur le revenu visent l'année précédente. On est par conséquent intéressés à obtenir des données sur le revenu des personnes qui formaient un ménage au moment de la collecte des données sur le travail en janvier, *indépendamment des déménagements* qui ont eu lieu entre janvier et mai. Cette dernière spécification a des répercussions sur les règles de suivi qui concernent les cohabitants. Les cas à considérer lors du premier cycle d'interviews pour la collecte des données sur le revenu sont les suivants :

- Pour les répondants longitudinaux âgés de 16 ans et plus qui ont déménagé entre les mois de janvier et mai, les règles sont les mêmes que celles qu'on a observées en janvier.

- C'est lors du premier cycle d'interviews sur le revenu qu'on a l'occasion de voir pour la première fois des *membres du ménage de retour après une absence* -- c'est-à-dire des répondants longitudinaux qui ont quitté le logement initial en 1993 et qui y sont revenus après la collecte des données sur le travail de janvier 1994. Cette catégorie de répondants très mobiles est une source de préoccupation dans les enquêtes longitudinales parce que si leur statut de membres du ménage de retour après une absence n'est pas signalé, ces répondants risquent d'être comptés deux fois (et même interviewés deux fois, dans certains cas). De plus, ils ne seraient pas traités comme des répondants longitudinaux et ne recevraient probablement pas le bon questionnaire. Pour éviter ces situations, chaque fois qu'il constate la présence d'un nouveau membre du ménage, l'intervieweur vérifie la liste des anciens membres du

ménage (qu'on appelle des *fantômes) pour s'assurer que la personne est bel et bien un nouveau membre et non un membre du ménage de retour après une absence. On recueille auprès des membres du ménage de retour après une absence les données sur le revenu, tandis qu'on réalise auprès des nouveaux membres du ménage une interview préliminaire.

- Les membres du ménage de retour après une absence sont notamment les répondants longitudinaux qui avaient déménagé dans un autre logement faisant partie du champ de l'enquête et qui ont réintégré leur logement initial. Ce sont également les personnes qui reviennent d'un séjour dans un établissement institutionnel ou à l'étranger. Ces personnes ont soit réintégré leur ancien logement, soit déménagé dans un nouveau. Dans les deux cas, la procédure à suivre est la même.
- Dans le cas des cohabitants (qui ne sont pas des nouveaux membres) âgés de 16 ans et plus qui ont déménagé entre janvier et mai et qui habitent toujours avec un répondant longitudinal, les règles sont les mêmes que celles qui s'appliquent aux répondants longitudinaux, c'est-à-dire qu'ils sont dépistés et interviewés dans le cadre de la collecte des données sur le revenu.
- Si un répondant longitudinal ou un cohabitant décédait entre janvier et mai, par délicatesse on ne recueille pas de données sur son revenu de l'année précédente.

- Les cohabitants âgés de 16 ans et plus qui n'habitent plus avec un répondant longitudinal sont aussi interviewés pour la collecte des données sur le revenu, mais ils ne le seront plus par la suite. En outre, les nouvelles personnes avec lesquelles ils vivent désormais ne sont pas interviewées du tout.
- Les nouveaux cohabitants (nouveaux membres) âgés de 15 ans et plus dont la présence est signalée pour la première fois en mai doivent répondre à l'interview préliminaire, mais non aux questions sur le revenu parce qu'ils ne vivaient pas avec un répondant longitudinal en janvier.
- On recueille les renseignements démographiques de base pour les nouveaux membres âgés de moins de 15 ans.

Deuxième cycle d'interviews sur le travail et au-delà

En janvier 1995, on a réalisé le deuxième cycle d'interviews pour la collecte des données sur le travail. C'est à ce moment que s'applique les dernières particularités des règles de suivi. Les cas à considérer sont les suivants :

- Les répondants longitudinaux âgés de 16 ans et plus qui ont déménagé dans un autre logement faisant partie du champ de l'enquête, dans les Territoires, dans un camp militaire, dans une réserve indienne ou aux États-unis sont dépistés à leur nouveau lieu de résidence et interviewés pour la collecte des données sur le travail, comme en janvier 1994.

- Les répondants longitudinaux âgés de 16 ans et plus qui sont des membres du ménage de retour après une absence sont interviewés au logement qu'ils occupent à ce moment pour la collecte des données sur le travail.
- Dans le cas d'un répondant longitudinal ou d'un cohabitant qui est décédé, on ne recueille pas d'autres renseignements que la date du décès. Du point de vue des opérations, la personne est éliminée de l'échantillon, mais les renseignements qui la concernent seront conservés dans le fichier de données.
- Pour les nouveaux membres âgés de 16 ans et plus, les procédures à suivre sont les mêmes qu'en janvier 1994.
- Les cohabitants qui n'habitent plus avec un répondant longitudinal sont interviewés pour cette année de collecte mais sont ensuite éliminés.
- On réalise auprès des cohabitants (y compris les nouveaux membres) et des répondants longitudinaux âgés de 15 ans une interview préliminaire, comme en janvier 1994.

Lors du troisième et des autres cycles d'interviews sur le travail, les procédures seront les mêmes que celles qui sont décrites ci-dessus. Quant aux procédures à suivre lors des collectes de données sur le revenu, elles seront les mêmes que pour le premier cycle.

2.5 Dépistage avec l'IAO

Comme on l'a mentionné précédemment, les personnes sélectionnées pour former l'échantillon de l'EDTR doivent être interviewées 13 fois : une interview préliminaire et deux interviews par année pendant six ans. Pour accomplir cette tâche avec succès, il est nécessaire de mettre en oeuvre des mécanismes permettant de dépister les répondants qui ont changé d'adresse. On a élaboré une stratégie générale relative au dépistage, laquelle comporte plusieurs éléments :

< *Renseignements recueillis au moment de la collecte des données*

On recueille dès le premier contact avec le répondant deux éléments d'information qui, au besoin, serviront pour le dépistage : premièrement le numéro de téléphone au travail et deuxièmement le nom, l'adresse et le numéro de téléphone d'un parent ou ami avec lequel on pourrait communiquer pour obtenir la nouvelle adresse du répondant. Ces renseignements seront mis à jour au besoin durant toute la période de l'enquête.

< *Dépistage effectué par l'intervieweur*

La plupart des données de l'EDTR seront recueillies au téléphone au cours d'une interview assistée par ordinateur (IAO). Si un intervieweur constate qu'un ménage a déménagé, le système de collecte automatisé présente différentes façons d'aborder le problème. L'une des possibilités consiste à utiliser le numéro de téléphone au travail du répondant ou à faire appel à la personne contact. Les autres moyens suggérés sont de recourir au service

d'assistance annuelle de la compagnie de téléphone ou d'utiliser l'information recueillie lorsqu'on a contacté l'ancienne adresse, etc.

< ***Dépistage au sein du bureau régional***

Si l'intervieweur ne peut résoudre le cas, il le transmet par modem au bureau régional désigné de Statistique Canada à l'aide du système IAO, qui garde en mémoire également quelles mesures ont déjà été prises de sorte qu'on ne refasse pas les mêmes démarches au bureau régional. Des *experts+ en dépistage du bureau régional ont accès aux annuaires par numéros et aux annuaires téléphoniques de toute leur région.

< ***Autres méthodes***

On examine actuellement d'autres possibilités (appariement avec un fichier de Poste Canada). Bien que de nombreuses options soient fonctionnelles, le rapport coût-efficacité représente le facteur déterminant dans la décision de les mettre en oeuvre.

Les résultats obtenus lors de la collecte de 1994 (janvier et mai) sont présentés au tableau 2.

TABLEAU 2

TAUX DE DÉPISTAGE	
Collecte 1994	
Janvier (travail)	Mai (revenu)
92,1%	80,7%

2.6 Taux de réponse de l'EDTR

Des taux de réponse sont calculés à chaque année. Cependant, les taux de réponse sont calculés de façon différente pour les taux de réponse transversaux et longitudinaux.

Taux de réponse transversal

Les taux de réponse transversaux sont définis au niveau du ménage (l'unité de collecte). Bien que pour des fins opérationnelles, des taux de réponse soient calculés pour chaque collecte, les taux de réponse transversaux, pour une année donnée, combinent l'information des deux collectes (janvier et mai) en un seul taux. En effet, comme les deux collectes se rapportent à une même période de référence et qu'elles sont combinées pour la diffusion du fichier de micro-données, il a été décidé de combiner les taux de réponse des deux collectes pour refléter un taux de réponse annuel.

Pour les calculs de taux de réponse transversaux, les ménages sont définis selon la composition du ménage de janvier (l'interview de mai est en général considérée comme une interview différée par rapport à celle de janvier). Un code de réponse pour l'interview de mai est donc recalculé basé sur la composition du ménage de janvier.

Pour le calcul du taux de réponse au niveau du ménage, on se base sur les codes de réponse des individus dans le ménage, que ce soit des personnes longitudinales ou des cohabitants. Un ménage répondant est défini comme un ménage ayant au moins une personne répondante. Une personne est définie répondante si elle a répondu soit à l'interview sur le travail ou à celle sur le revenu. Finalement, une

interview est considérée comme répondue si un minimum de questions clés ont été répondues.

Les ménages répondants sont divisés en ménages complètement répondants et en ménages partiellement répondants. Les ménages partiellement répondants sont pondérés. Les données manquantes dans ces ménages sont imputées (un minimum de champs sont imputés, les autres sont définis comme manquants).

Taux de réponse longitudinal

Les taux de réponse longitudinaux sont quant à eux définis au niveau de l'individu. En effet, il est difficile de définir le ménage comme unité d'analyse longitudinale.

Le taux de réponse longitudinal est basé seulement sur l'échantillon longitudinal, même s'il se peut que l'on possède de l'information longitudinale pour certains cohabitants (en effet un cohabitant est interviewé dans la mesure où il demeure avec la personne longitudinale).

Pour le taux de réponse longitudinal, on considère un individu répondant s'il a répondu à toutes les années d'enquête (pour une année donnée, la personne est définie comme répondante si elle a répondu à l'entrevue sur le travail ou à celle sur le revenu). On évalue présentement la possibilité d'inclure comme répondants les individus qui n'auraient qu'une année de données manquantes, dans la mesure où l'année manquante est bornée par deux années de réponse.

Les taux de réponse transversaux et longitudinaux obtenus lors de la collecte de 1993 (interview préliminaire) et de la collecte de la collecte de 1994 (interview sur le travail et le revenu) sont présentés au tableau 3. Dès la première année, le taux

de réponse transversal diffère du taux de réponse longitudinal puisque l'échantillon de l'EDTR a été choisi parmi l'échantillon de l'EPA.

TABLEAU 3

TAUX DE RÉPONSE			
Collecte 1993 (préliminaire)		Collecte 1994 (travail et revenu)	
transversal	longitudinal	transversal	longitudinal
%	%	%	%
88,5	84,8	90,9	77,0

2.7 Imputation

Pour plusieurs enquêtes sociales, on compense pour la non-réponse totale par la pondération, alors que la non-réponse partielle est souvent imputée. Puisque l'EDTR est une enquête longitudinale, la dimension temps ajoute à la complexité de l'élaboration d'une stratégie. Différents points doivent être considérés afin d'établir une stratégie d'imputation, par exemple, l'impact des collectes différées. L'EDTR fait deux collectes par année qui se reportent à une même période de référence. De plus, ces deux collectes sont combinées pour ne produire qu'un seul fichier de micro-données à chaque année. Il est donc tentant de penser à une vague de collecte manquante comme à une non-réponse pour un bloc d'items et d'imputer pour les valeurs manquantes. Pour déterminer ce que l'on doit imputer, les règles générales suivantes ont été établies :

- À chaque année, pour une personne admissible à l'enquête, on impute si la personne a répondu seulement à une des deux phases

(soit travail ou soit revenu). Cette condition est vraie pour les personnes longitudinales et pour les cohabitants.

- À chaque année, pour une personne admissible à l'enquête, on impute si la personne est non-répondante aux deux phases (travail et revenu) et que son ménage est partiellement répondant (au moins un membre de son ménage a répondu). Le ménage est défini en janvier de chaque année.
- Pour une personne répondante, on impute si un item est manquant ou il y a incohérence.

Cependant, parce que l'EDTR est une nouvelle enquête longitudinale et que c'est la première année de collecte, on limite un peu l'imputation pour l'année de référence de 1993. Seules les variables démographiques nécessaires à la pondération telles province, âge et sexe, et les variables sur le revenu sont imputées. Les variables de l'interview sur le travail ne sont pas imputées pour l'année de référence 1993. Le manque de temps et d'informations, pour imputer adéquatement, ont aidé à prendre cette décision. Toutefois, les variables sur le revenu sont imputées et la stratégie est la suivante :

- Imputation par **le plus proche voisin** pour la majorité des variables : salaires et traitements, revenus nets d'un emploi autonome agricole, revenus nets d'un emploi autonome non agricole, investissement, gains en capital, prestations d'assurance-chômage, assistance sociale et suppléments de revenu provinciaux, indemnités pour accidents de travail, pension de retraite et rentes,

retraits d'un REER, pensions alimentaires, autres revenus monétaires imposables.

- Imputation à l'aide d'un **modèle de régression par échelon** pour les variables "*impôt fédéral et provincial sur le revenu versé*". Les variables utilisées dans le modèle pour l'impôt fédéral sont : revenu total imposable, classe de travailleur, relation, état matrimonial, nombre de dépendants et occupations. Pour l'impôt provincial, on a utilisé uniquement la variable impôt fédéral.
- Imputation **déterministe** pour les variables : prestations fiscales pour enfants, crédit pour la taxe sur les produits et services et supplément de revenu garanti et allocation du conjoint.

Lors des vagues suivantes, on pourra considérer l'imputation longitudinale. Celle-ci consiste à utiliser les données de la vague précédente pour imputer les données manquantes ou/et incohérentes de la vague courante.

2.8 Pondération

L'EDTR est d'abord et avant tout une enquête longitudinale. Une stratégie a donc été développée pour traiter de la pondération longitudinale. Cependant, bien que le but premier soit longitudinal, il existe toujours une certaine demande pour pouvoir produire des estimations transversales. En conséquence, à chaque année, une pondération longitudinale et une pondération transversale sont donc effectuées.

Pondération longitudinale

Une pondération longitudinale est faite de façon indépendante pour chaque panel. Éventuellement, on voudrait créer un fichier longitudinal après trois ans et après six ans. Il y aurait donc deux pondérations longitudinales pour un panel. Cependant, pour le premier panel, une pondération longitudinale sera faite à chaque année (il y aura donc six fichiers longitudinaux, avec six poids longitudinaux).

Un panel longitudinal n'est pas "mis à jour" après sa sélection ; il est donc représentatif de la population, au moment de sa sélection. Pour le premier panel, l'échantillon longitudinal a été sélectionné (et sera donc pondéré) pour représenter la population des 10 provinces canadiennes, en janvier 1993, qui faisait partie du champ de l'enquête. À chaque année, la pondération est donc ajustée à cette population.

Avant d'effectuer la pondération, des critères déterminent si la personne est éligible à la pondération longitudinale. Une personne est éligible à la pondération longitudinale si elle fait partie de l'échantillon longitudinal sélectionné pour ce panel. Les enfants et les personnes qui déménagent à l'extérieur des 10 provinces, ou qui entrent en institution sont éligibles à la pondération, bien que des données d'enquête ne soient pas recueillies pour eux.

L'échantillon longitudinal a été sélectionné à partir de l'échantillon de l'EPA. Au moment de l'entrevue de l'EPA, les répondants ont été informés qu'ils étaient choisis pour faire partie de l'EDTR et quelques questions supplémentaires leur ont été posées (on réfère souvent à ce premier contact comme étant l'entrevue préliminaire). L'EPA a un taux de réponse de 95%. De plus, 88% des répondants

à l'EPA ont accepté de participer à l'entrevue préliminaire. À cause de contraintes budgétaires, seulement 84% des répondants à l'entrevue préliminaire ont été interviewés par la suite. Un petit panel de répondants à l'EPA mais non-répondants à l'entrevue préliminaire (environ 200 ménages) a été suivi. Il servira à des études pour fins de qualité cependant, il ne sera pas pondéré.

Pour la pondération longitudinale, on possède beaucoup d'information sur les non-répondants (la seule exception est pour les gens qui n'ont pas participé à l'EPA et un ajustement pour la non-réponse est déjà fait par l'EPA pour compenser pour cette non-réponse).

Différentes analyses montrent que les caractéristiques connues entre les répondants et les non-répondants sont différentes; on retrouve plus de répondants parmi les personnes employées, parmi les gens qui n'ont pas déménagé, etc.

Il a été démontré que la performance de l'estimateur longitudinal est améliorée dans la mesure où le facteur de compensation pour la non-réponse explique le comportement de réponse. L'ajustement pour la non-réponse se base donc sur des groupes de réponse homogènes. Les variables qui forment ces groupes de réponse sont déterminées à l'aide de la régression logistique et de modèles de détections d'interaction. Le but de ces classes d'ajustement est de tenter de compenser pour un certain biais qui serait introduit par le fait que le processus de réponse n'est pas uniforme.

Après la correction pour la non-réponse, une post-stratification est faite par province, groupe d'âge et sexe (comme mentionné plus haut, les post-strates sont définies par rapport à la période de référence.)

Pondération transversale

La pondération transversale vise à produire des estimations qui sont représentatives de la population pour une année donnée. Pour ce faire, on a besoin de pouvoir mettre à jour l'échantillon pour tenir compte de nouvelles personnes qui font partie de la population.

La plus importante source de mise à jour de la population est l'introduction d'un nouveau panel à tous les trois ans. À ce moment, l'échantillon transversal va combiner les données de deux panels et donc doubler la taille de l'échantillon transversal.

Il existe une autre source de "mise à jour"; à chaque année, les personnes de l'échantillon longitudinal qui déménagent sont dépistées. Parce qu'on s'intéresse aux "variables" familiales de l'individu, on interroge aussi les personnes qui cohabitent avec la personne longitudinale. Ces cohabitants peuvent faire partie de la population originale (donc elles avaient une probabilité de sélection différente de zéro) mais elles n'ont pas été échantillonnées. Ils peuvent aussi être de "nouveaux arrivants" dans la population. Une technique appelée le partage des poids (développée par Ernst [2]) permet d'inclure dans la pondération transversale les cohabitants, en se servant des probabilités de sélection de l'échantillon longitudinal. De la recherche effectuée par Lavallée [3], il a été démontré que l'estimateur ainsi construit est non biaisé.

La pondération transversale se fait sur les individus dans les logements avec au moins une personne longitudinale, en janvier de chaque année (le logement doit être dans une des dix provinces canadiennes). Suite à l'ajustement de non-réponse fait lors de la pondération longitudinale, on effectue le partage des poids.

Finalement une post-stratification est faite par province, sexe et groupe d'âge (de l'année courante).

2.9 Confidentialité

L'EDTR diffusera des fichiers de microdonnées. Ces fichiers doivent donner les résultats de l'enquête de façon à ce qu'on ne puisse pas identifier un répondant. L'équipe de l'EDTR élabore une stratégie pour rendre négligeable le risque d'appariement avec d'autres fichiers administratifs et ce, afin de dissuader des pirates à utiliser ce fichier pour identifier des individus et des familles, ou d'obtenir de l'information supplémentaire sur ces derniers. Les méthodes suivantes seront utilisées : les suppressions (identification de la personne,...), les regroupements de modalités (année de naissance), les arrondissement non aléatoires (variables de revenu), l'imputation (variables de revenu) et les arrondissements aléatoires (variables de revenu).

3. Conclusion

Présentement, différentes études sont entreprises concernant l'évaluation de la qualité de nos données. En raison de leur complexité, les enquêtes par panel posent certaines difficultés intéressantes pour la mesure de la qualité des données. Les normes et les paramètres établis en matière d'évaluation de la qualité visent en général des enquêtes transversales et doivent faire l'objet d'une adaptation. Un important programme de mesure de la qualité des données est actuellement en cours d'élaboration pour l'EDTR. Les études sur la qualité des données seront élaborées de façon à évaluer la réponse et les erreurs de réponse ainsi que les répercussions de l'interview assistée par ordinateur et de l'interview avec rétro-information. Ces études permettront de comparer les résultats avec des données

de référence tirées de dossiers fiscaux, de l'Enquête sur la population active et d'autres sources.

Les premières microdonnées seront prêtes en 1995. Ce fichier contiendra les données de l'interview préliminaire, réalisée en janvier 1993, et les renseignements recueillis lors de la première vague d'interviews sur le travail et le revenu en janvier et en mai 1994. Au moment de la diffusion du fichier de microdonnées, on mettra à la disposition des utilisateurs un document d'accompagnement afin de les aider à utiliser les données. Chaque année, on diffusera un nouveau fichier de microdonnées, qui contiendra toutes les données recueillies à ce jour et remplacera le fichier de l'année précédente.

Bibliographie

Les travaux d'élaboration de l'enquête ont fait l'objet de nombreuses publications. On peut se procurer gratuitement le bulletin trimestriel *La Dynamique* en communiquant avec l'Unité de diffusion, Édifice Jean-Talon, 7^e étage, section B5, Statistique Canada, Ottawa (Ontario), Canada, K1A 0T6, par INTERNET: DYNAMIQUE@STATCAN.CA; par téléphone au (613) 951-7355 ou sans frais au 1-800-297-7355; ou par télécopieur au (613) 951-3012. Il existe également une série complète de documents de travail qui décrivent diverses caractéristiques, certains aspects et les décisions qui ont été prises relativement à l'EDTR depuis 1992. La série de documents de recherche est offerte, sur papier ou disquette, à un prix modéré. Chaque numéro de *La Dynamique* contient un résumé des documents de travail récemment publiés.

- [1] SINGH, M.P. et al., *Méthodologie de l'enquête sur la population active du Canada 1984-1990*. Publication de Statistique Canada. Numéro de Catalogue 71-526.

- [2] ERNST, L.R. (1989). *Weighting issues for longitudinal household and family estimates*. Dans *Panel Surveys*, (Éds. D. Kasprzyk, G. Duncan, G. Kalton et M.P. Singh), New York: John Wiley, 139-159.

- [3] LAVALLÉE, P. *Pondération transversale des enquêtes longitudinales menées auprès des individus et des ménages à l'aide de la méthode du partage des poids*. *Techniques d'enquête*, (Éds. par Statistique Canada), Volume 21, numéro 1, juin 1995, 27-35.