

N° de catalogue 94-16

**LA MESURE DE LA NON-RÉPONSE DANS UNE
ENQUÊTE LONGITUDINALE : L'EXPÉRIENCE DE
L'ENQUÊTE SUR LA DYNAMIQUE DU TRAVAIL
ET DU REVENU**

Septembre 1994

Sylvie Michaud, Division des méthodes d'enquêtes sociales

Maryanne Webber, Division des enquêtes-ménages

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

Ce document a été présenté au cinquième atelier international sur la non-réponse aux enquêtes-ménages, lequel a eu lieu à Ottawa en septembre 1994.

Les taux de réponse et l'estimation de la variance sont des éléments d'évaluation clés de la qualité d'une enquête. L'Enquête sur la dynamique du travail et du revenu est une nouvelle enquête longitudinale qui a été mise en oeuvre en 1994. Comme il s'agit d'une enquête longitudinale et que les données sont recueillies au moyen de l'interview assistée par ordinateur, nous avons besoin de mesures particulières pour pouvoir évaluer la qualité des données. Même le plan de sondage a ses particularités, comme l'interview différée sur le revenu et le recours à l'interview avec rétro-information. Dans ce document, il est question des diverses mesures de qualité des données qui sont proposées pour les besoins de l'enquête, incluant un taux de réponse de la vague, un taux de réponse du panel et la détermination des caractéristiques des non-répondants. Un éventail de mesures du genre est nécessaire pour pouvoir comprendre les divers aspects du mécanisme de réponse.

TABLE DES MATIÈRES

	Page
1. INTRODUCTION	1
2. DESCRIPTION DE L'EDTR	2
3. MESURES DE QUALITÉ DES DONNÉES	5
4. TAUX DE RÉPONSE ET DE NON-RÉPONSE	5
4.1 DÉFINITION DES TERMES LIÉS AUX TAUX DE RÉPONSE	6
4.2 INCIDENCE DU PLAN DE L'EDTR SUR LA DÉFINITION DES TAUX DE RÉPONSE	7
4.3 TAUX DE RÉPONSE PROPOSÉS POUR L'EDTR	10
5. DIFFÉRENCES ENTRE LES RÉPONDANTS ET LES NON-RÉPONDANTS	16
5.1 ANALYSE DES CARACTÉRISTIQUES DES NON-RÉPONDANTS FONDÉE SUR LA RÉPONSE À DES VAGUES PRÉCÉDENTES	16
5.2 COUPLAGE STATISTIQUE DES DONNÉES DE L'ÉCHANTILLON AVEC LES DONNÉES FISCALES	17
5.3 SUIVI D'UN ÉCHANTILLON DE NON-RÉPONDANTS DE L'INTERVIEW PRÉLIMINAIRE	19
6. CONCLUSIONS	20
RÉFÉRENCES	21
ANNEXE 1 DÉFINITION DES UNITÉS AUX FINS DE CALCUL DES TAUX DE RÉPONSE	22

1. INTRODUCTION

La mesure de l'erreur due à la non-réponse comporte deux aspects : le taux de non-réponse et l'évaluation des différences entre répondants et non-répondants. Il est relativement facile de mesurer le taux de non-réponse à une enquête et il s'agit souvent d'une des principales mesures de qualité de données. Il est dans la plupart des cas plus difficile de cerner les différences entre répondants et non-répondants en raison du peu des renseignements dont nous disposons sur les non-répondants.

Statistique Canada a élaboré des normes en matière d'établissement de taux de réponse dont l'utilité ne fait aucun doute. Cependant, ces normes ont avant tout été conçues en fonction d'enquêtes transversales. Comme plusieurs nouvelles enquêtes longitudinales sur des sujets à caractère social sont en voie d'élaboration, le moment est sans doute tout indiqué pour se demander quelles mesures supplémentaires il conviendrait de prendre de façon à tracer un portrait fidèle et uniforme de la qualité des données des enquêtes longitudinales. Souvent, lorsque le contact avec les répondants est rompu après la première interview d'une enquête longitudinale, on dispose de bon nombre de renseignements sur les caractéristiques de ces non-répondants et une telle information peut être prise en compte dans la mesure de la non-réponse totale.

L'Enquête sur la dynamique du travail et du revenu (EDTR) est l'une de ces nouvelles enquêtes longitudinales. Ce document présente les problèmes que liés à la définition des taux de réponse et d'autres mesures nécessaires pour permettre aux utilisateurs d'évaluer la qualité des données.

2. DESCRIPTION DE L'EDTR

Une brève vue d'ensemble de l'Enquête sur la dynamique du travail et du revenu est fournie, de façon à situer dans son contexte l'analyse de la mesure de la non-réponse.

Caractéristiques de l'enquête L'EDTR est une enquête menée auprès de ménages. L'échantillon du premier panel a été tiré de l'échantillon de l'Enquête sur la population active (EPA). L'échantillon de l'EPA est un échantillon aréolaire probabiliste à plusieurs degrés, couvrant la population de toutes les provinces, à l'exclusion des personnes vivant dans une réserve indienne, des militaires et des pensionnaires d'un établissement institutionnel. Ces mêmes critères de sélection ont été retenus pour l'EDTR (à l'étape de la sélection de l'échantillon initial), à l'exception d'un seul : les membres des Forces armées vivant en dehors des casernements sont compris dans le champ de l'enquête.

Un panel dure six ans (en excluant l'entrevue préliminaire). À la fin de cette période, treize interviews auront été réalisées auprès des personnes faisant partie de ce panel : soit une interview préliminaire, au moment de la sélection de l'échantillon, dans le but de recueillir les données de base, plus six interviews sur le travail (chaque année en janvier) et six interviews sur le revenu (chaque année en mai). Les interviews sur le travail et sur le revenu portent dans les deux cas sur l'année civile précédente.

Les panels de l'EDTR se chevauchent. Le premier panel a été sélectionné en janvier 1993, et à ce moment on a fait l'interview préliminaire. (Par conséquent, la première vague d'interviews sur le travail et le revenu a débuté en 1994.) Un deuxième panel sera introduit dans l'échantillon en 1996, de sorte que l'échantillon transversal deviendra complet à ce stade de l'enquête. Au moment de

l'introduction du troisième panel en 1999, le premier panel sera retiré de l'échantillon.

Au départ, le premier panel comptait 15 000 ménages. Ce total comprend environ 31 000 personnes âgées de 16 ans et plus qui répondront à des questions au sujet du travail et du revenu. La taille du deuxième panel n'a pas encore été déterminée.

Règles de suivi Tous les membres des ménages faisant partie des logements sélectionnés pour faire partie de l'échantillon longitudinal doivent être suivis pendant six ans, qu'ils déménagent ou qu'ils forment deux ou plusieurs autres ménages distincts. Les *répondants longitudinaux* sont les personnes qui font partie de l'échantillon depuis le début. Les interviews sont aussi réalisées auprès des personnes qui emménagent avec un répondant longitudinal au cours des six années d'enquête (ces personnes sont appelées *cohabitants*), tant que ces nouveaux membres vivent avec un répondant longitudinal.

Le suivi se fait auprès de tous les répondants, quel que soit leur âge. Les interviews sur le travail et le revenu ne s'appliquent qu'aux répondants longitudinaux et cohabitants âgés de 16 ans et plus le 1^{er} janvier de l'année d'enquête. Cependant, un répondant plus jeune devient éligible aux interviews lorsqu'il atteint l'âge de 16 ans. L'année de ses 15 ans, il devient éligible pour l'interview préliminaire, en préparation des interviews sur le travail et le revenu subséquentes.

L'interview de mai sur le revenu ne constitue pas une vague distincte, mais plutôt une interview «différée» par rapport à l'interview de janvier. En effet, on s'est rendu compte que les données sur le revenu étaient de meilleure qualité lorsque les répondants pouvaient consulter leur déclaration de revenus. Le fait que les interviews de janvier et de mai s'inscrivent dans la même vague du point de vue

conceptuel influe sur les règles de suivi. Ainsi, tout cohabitant qui cesse d'habiter avec le répondant longitudinal entre les mois de janvier et mai fait quand même l'objet d'un dépistage et d'une interview en mai, afin d'obtenir les données relatives à son revenu (de l'année précédente). Il faut donc attendre ce moment pour retirer cette personne de l'échantillon.

Caractéristiques opérationnelles Les données de l'EDTR sont recueillies par l'interview assistée par ordinateur (IAO). La collecte des données est décentralisée; en effet, les intervieweurs travaillent habituellement chez eux, menant l'interview par téléphone et transmettant les données par modem au bureau régional le plus proche. La déclaration par personne interposée est acceptée, dans la mesure où cette autre personne, le «répondant-substitut» connaît bien l'enquêté. Autrement, l'intervieweur doit convenir d'un moment pour rappeler.

L'EDTR a été conçue de manière à ce qu'on puisse tirer profit des flexibilités de l'IAO susceptibles d'améliorer la qualité des données. Ainsi, l'IAO permet la vérification interactive des dates déclarées par le répondant. L'IAO facilite également l'interview avec rétro-information, technique qui consiste à rappeler aux répondants des renseignements fournis lors d'interviews précédentes afin de les aider à mieux se rappeler certaines situations et de réduire ainsi les problèmes de lisibilité et le nombre de fausses transitions.

3. MESURES DE QUALITÉ DES DONNÉES

Comme nous l'avons déjà souligné, la mesure de l'erreur due à la non-réponse comporte deux aspects : la détermination du taux de non-réponse et l'évaluation des différences entre les répondants et les non-répondants. Groves [1] note que le taux de non-réponse peut à son tour être réparti en diverses catégories, le comportement pouvant varier selon le genre de non-réponse (par ex. refus, personnes non dépistées, autres cas de non-réponse).

Dans une enquête longitudinale, il est particulièrement important de pouvoir déterminer le genre de non-réponse compte tenu du fait que les facteurs à l'origine de la non-réponse peuvent changer en cours de route du panel. Habituellement, la non-réponse attribuable aux refus est plus élevée en début de panel et a tendance à diminuer par la suite. Par contre, lors de la première interview, le non-dépistage n'entre pas dans les causes de non-réponse. L'effet de ce facteur peut croître de façon considérable au cours des vagues subséquentes.

Des mesures d'assurance de la qualité des données ont été prévues au moment de la conception de l'EDTR. Toutefois, certaines des caractéristiques du plan de l'enquête rendent plus complexe la tâche de détermination des taux de réponse. En même temps, certains aspects de l'enquête facilitent l'évaluation des différences entre les répondants et les non-répondants.

4. TAUX DE RÉPONSE ET DE NON-RÉPONSE

Le taux de non-réponse constitue souvent une mesure clé de la qualité des données. Statistique Canada a élaboré des normes pour l'évaluation des taux de réponse [2]. On ne peut calculer des taux de réponse sans d'abord s'entendre sur la

façon de définir les termes. Ces définitions influent grandement sur la mesure du taux de réponse.

4.1 DÉFINITION DES TERMES LIÉS AUX TAUX DE RÉPONSE

Les diverses unités qui entrent dans le calcul des taux de réponse, fondées sur les normes élaborées par Statistique Canada, sont les suivantes (pour une brève description, se reporter à l'annexe 1) :

Nombre total d'unités	(1)	= (2) + (3)
Cas résolus	(2)	= (4) + (12)
Cas non résolus	(3)	
Unités dans le champ de l'enquête	(4)	= (5) + (8)
Unités répondantes	(5)	= (6) + (7)
Réponses après refus	(6)	
Autres répondants	(7)	
Unités non répondantes	(8)	= (9) + (10) + (11)
Refus	(9)	
Unités non contactées	(10)	
Unités de non-réponse résiduelles	(11)	= (16) + (17) + (18) (nouveau)
Unités hors du champ de l'enquête	(12)	= (13) + (14) + (15)
Unités inexistantes	(13)	
Unités provisoirement hors du champ de l'enquête	(14)	
Unités hors du champ de l'enquête en permanence	(15)	

Les unités de non-réponse résiduelles (11) devraient être subdivisées en trois catégories :

Unités non dépistées	(16)
Données non reçues (problèmes de transmission)	(17)
Autres cas de non-réponse	(18)

Deux types de taux de réponse sont proposés : un taux de réponse opérationnel et un taux de réponse à des fins de pondération.

Le taux de réponse opérationnel (TR OP) est défini comme le rapport du nombre de répondants (interviews complètes) sur le nombre de cas résolus éligibles à l'enquête ou les cas non résolus. Dans le taux de réponse utilisé à des fins de pondération (TR POND), le nombre de cas non résolus est réparti entre un nombre estimé d'unités éligibles et un nombre estimé d'unités non éligibles. Le nombre estimé d'unités non éligibles est exclu du calcul du taux de réponse.

Dans le calcul des taux, l'unité répondante peut être le logement ou la personne, même si le logement est l'unité la plus utilisée dans un grand nombre d'enquêtes sociales.

4.2 INCIDENCE DU PLAN DE L'EDTR SUR LA DÉFINITION DES TAUX DE RÉPONSE

La première difficulté est celle que pose l'interview différée sur le revenu. Les interviews de janvier et de mai forment une «vague». Un seul fichier de données sera constitué chaque année, combinant l'information recueillie au cours des deux périodes de collecte. Pour répartir les unités éligibles en répondants et non-répondants, la non-réponse peut être définie de deux façons. Tout d'abord, les

répondants peuvent être restreints aux personnes qui ont répondu aux questions des deux interviews. Cette définition entraîne une perte substantielle d'information. L'autre solution (qui maximise le taux de réponse) définit comme répondants les personnes ayant répondu aux questions d'au moins une des interviews de janvier et mai. Cette définition implique un taux plus élevé de non-réponse par item.

La deuxième difficulté découle du fait que l'EDTR interviewe tous les membres du ménage. Le schéma de pondération transversale de l'enquête repose sur une méthode de pondération intégrée qui consiste à rajuster les coefficients de pondération en fonction de la composition de la famille, de façon à donner le même poids à tous les membres du ménage. Le problème se pose dans le cas des ménages partiellement répondants. On peut les exclure complètement ou les inclure en ayant recours à l'imputation pour le membre du ménage manquant. (Soulignons que l'imputation peut être aussi simple que d'inscrire «ne sait pas» comme réponse à toutes les questions concernant le non-répondant.)

L'interview différée complique encore une fois les choses : les ménages sont dynamiques et leur composition peut changer entre les mois de janvier et mai. De plus, comme l'enquête suit les répondants longitudinaux, c'est-à-dire toutes les personnes qui étaient membres du ménage initial, après la première année, ce n'est qu'au moment de la collecte qu'on peut déterminer le nombre exact de ménages.

La troisième difficulté est liée à l'utilisation de l'interview assistée par ordinateur (IAO). L'IAO comporte de nombreux avantages, mais il s'agit d'une technologie d'utilisation récente à Statistique Canada et il y a encore beaucoup à apprendre. Comme le montre l'exemple qui suit, l'IAO influe sur la réponse et ajoute une nouvelle complexité à sa mesure.

Comme les IAO sont décentralisées, des problèmes de transmission peuvent se poser. Ces derniers se répartissent en deux catégories et devraient donc être traités de manière différente dans la recherche d'explications du mécanisme de réponse. Dans la première catégorie se trouvent les cas que l'intervieweur n'a jamais reçus, ou cas «non démarrés». Ces cas peuvent probablement être traités comme suivant un mécanisme de non-réponse aléatoire.

La deuxième catégorie englobe les cas reçus par l'intervieweur, mais non reçus au bureau central. Contrairement aux cas «non démarrés», il s'agit généralement de ménages non répondants perdus au cours d'une des opérations de transmission. Certaines de ces opérations sont complexes et font appel à plusieurs transferts de données. Par exemple, des cas de refus doivent être référés, et donc transmis, à un intervieweur principal, puis au gérant de projet. De même, le dépistage se fait à deux niveaux; les personnes non dépistées par l'intervieweur sont transmises à un «dépisteur» au bureau régional. Parce que les cas n'ont jamais été reçus au bureau central, il est difficile de savoir exactement ce qui s'est passé, et les raisons pour la non réponse.

L'attribution de codes de réponse justes est une autre difficulté associée à l'IAO. Pour aider l'intervieweur, certains codes de réponse sont attribués automatiquement, en particulier pour les entrevues complètes ou si la personne n'est pas éligible à l'interview (par exemple, la personne a quitté le ménage pour aller séjourner pendant plus de six mois dans une institution). Les autres codes de réponse -- réponse partielle, personne non dépistée et autres cas -- sont attribués manuellement par l'intervieweur. Il y a deux types de problèmes. L'intervieweur pourrait éventuellement inscrire «refus» ou «ne sait pas» chaque fois qu'une réponse doit être entrée, mais comme le système assigne automatiquement un code de réponse «complet» après la dernière question, ce dernier est incorrect. De plus, le choix d'un code de réponse est souvent subjectif. Par exemple, si la définition

d'«interview partielle» n'est pas assez claire, un certain nombre de cas classés «partiels» doivent être recodés et comptabilisés dans la non-réponse parce que l'information fournie n'est pas suffisante.

4.3 TAUX DE RÉPONSE PROPOSÉS POUR L'EDTR

Compte tenu des caractéristiques particulières de l'EDTR, nous pensons que les deux types de taux proposés dans les standards ne suffisent pas. Trois types de taux de réponse sont proposés : un taux de réponse de phase, un taux de réponse de vague et un taux de réponse longitudinal.

4.3.1 Taux de réponse de phase

Le taux de réponse de phase est un taux transversal. Il sert à évaluer la qualité d'une période de collecte précise. Par conséquent, pour l'EDTR, il sera calculé deux fois par année : en janvier et en mai. Ce taux sera calculé au niveau des ménages; il pourrait aussi être calculé au niveau des personnes, mais uniquement pour le sous-ensemble des répondants longitudinaux. En effet, le nombre des cohabitants non répondants est sous-estimé, car ces derniers ne sont pas nécessairement énumérés dans les ménages non répondants.

Le tableau 1 montre la répartition des ménages de l'échantillon selon le code de réponse de phase janvier. Comme on peut le voir, les cas non reçus en raison d'un problème de transmission ont été la cause la plus importante de non-réponse. À partir des données recueillies en mai, nous pouvons estimer qu'environ 5 % des cohabitants (en majorité des non-répondants) n'ont pas été inclus dans les non réponses de janvier.

Selon les standards actuels, les réponses partielles doivent être classés en répondants ou non-répondants. On prévoit les classer comme répondants et procéder à une imputation pour les questions principales pour la personne non répondante du ménage. Selon cette définition, le taux de réponse est de 87 %.

Tableau 1 Répartition de l'échantillon des ménages interviewés en janvier

	Nombre de ménages
Ménages initiaux	15 006
Ménages créés	1 024
Nombre total prévu de ménages	16 031
Non éligibles	138
Réponse (totale)	13 113
Réponse (partielle)	708
Non-réponse - refus	276
Non-réponse - personnes non dépistées	334
Non-réponse - cas non reçus	1 156
Non-réponse - autres cas	306

4.3.2 Taux de réponse de vague

Le taux de réponse de vague fournit une vue d'ensemble de ce qui s'est passé au cours d'une année de collecte. Un seul fichier de données est produit annuellement combinant les interviews distinctes du travail et du revenu. Le taux de réponse de vague est donc une mesure pour les utilisateurs des données. Ce taux sera établi pour les ménages comme pour les individus, pour des raisons différents.

À cause de l'interview différée sur le revenu, le taux de réponse de vague peut être calculé de deux façons. Le numérateur peut se limiter à inclure les personnes ayant répondu aux deux interviews (seulement). Alternativement, il peut inclure toutes les personnes ayant répondu à *au moins une* des interviews.

Le tableau 2 présente la répartition de l'échantillon des répondants longitudinaux éligibles (c'est-à-dire les personnes âgées de 16 ans et plus) et les résultats selon les deux modes de calcul. Dans les colonnes 1 et 2, on peut voir la répartition des codes de réponse déterminés automatiquement par le logiciel d'application. La colonne 3 montre le résultat après recodage (effectué principalement lorsqu'il n'y avait pas d'information suffisante). Enfin, la colonne 4 montre le taux de réponse de vague qui combine les deux interviews. Pour les cas non reçus en janvier à cause d'un problème de transmission, mais reçus en mai, le code de non réponse de mai a été imputé à janvier.

Tableau 2 Codes de réponse de vague de l'échantillon longitudinal
(personnes éligibles) après la première année d'enquête

	Travail	Revenu	Revenu-après recodage	EDTR- interviews combinées
Réponse totale	25 792 (84,7 %)	24 101 (79,1 %)	23 003 (75,6 %)	21 328 (70,0 %)
Travail seulement	-	-	-	4 464 (14,7 %)
Revenu seulement	-	-	-	1 675 (5,5 %)
Refus	598 (1,9 %)	2 246 (7,4 %)	3 694 (12,1 %)	1 194 (3,9 %)
Personnes non dépistées	312 (1,0 %)	550 (1,8 %)	546 (1,8 %)	530 (1,7 %)
Autres	1 071 (3,5 %)	2 354 (7,7 %)	2 153 (7,1 %)	683 (2,2 %)
Cas non reçus- problème de transmission	2 435 (8,0 %)	931 (3,1 %)	786 (2,6 %)	308 (1,0 %)
Personnes devenues non éligibles	253 (0,8 %)	279 (0,9 %)	279 (0,9 %)	279 (0,9 %)
Total	30 461 (100 %)	30 461 (100 %)	30 461 (100 %)	30 461 (100 %)

Comme on peut le voir dans la colonne 4 du tableau 2, le taux de réponse de la première vague est de 71 % lorsque la première définition est utilisée et il grimpe à

91 % lorsque nous appliquons la seconde définition. Il est clair que cette deuxième définition maximise le taux de réponse, mais en même temps elle augmente le taux de non réponse d'items. Les causes de non-réponse sont différentes entre l'interview sur le travail et celle sur le revenu. Le recodage des codes de non-réponse a augmenté le taux de personnes non dépistées et les refus. Le recodage des codes de réponse augmente considérablement le nombre de refus à l'interview sur le revenu.

L'EDTR a l'intention d'utiliser la seconde définition de taux de réponse. Des méthodes d'imputation sont développées pour les questions clés de l'interview manquante. Les répondants seront classés en deux catégories : ceux qui ont participé aux deux interviews et ceux qui n'ont répondu qu'à une des deux. Le taux de réponse de vague calculé au niveau des personnes indiquera aux utilisateurs le nombre de personnes pondérées, à des fins d'analyse transversale.

Il est plus difficile de calculer le taux de réponse de vague au niveau des ménages. En effet, les ménages ont une structure dynamique, mais pour les besoins de l'établissement d'un taux, ils doivent être définis à un moment précis dans le temps (janvier dans le cas qui nous intéresse).

Le tableau 2 est un tableau simplifié de l'ensemble des résultats et des codes. Par exemple, certaines personnes sont devenues inéligibles entre les mois de janvier et mai. Ces cas doivent être recodés et comptés comme non-répondants aux fins du calcul des taux de réponse, compte tenu du fait que les critères d'éligibilité sont basés sur la situation en janvier. En janvier, les personnes âgées de 15 ans répondent aussi à une interview préliminaire. Par conséquent, au niveau des personnes, un taux distinct devrait être établi pour les répondants âgés de 15 ans.

4.3.3 Taux de réponse longitudinal

Un taux de réponse longitudinal sera établi uniquement au niveau des personnes, en raison de la difficulté que pose la définition de ménages longitudinaux. Il s'agit d'un taux de réponse cumulatif qui servira d'indicateur de l'érosion de l'échantillon, c'est-à-dire essentiellement une mesure de réponse par rapport à l'échantillon initial.

Le premier contact avec les répondants de l'échantillon longitudinal de l'EDTR a eu lieu un an avant que ne débute la première vague d'interviews sur le travail et le revenu. Ce premier contact correspond à ce que l'on a appelé l'interview préliminaire. Le taux de réponse à l'interview préliminaire des membres du premier panel de l'EDTR était de 88 %. Le taux de réponse longitudinal est un rapport qui exprimera le nombre de répondants longitudinaux de chacun des vagues successifs en proportion du nombre de personnes ayant participé à l'interview préliminaire.

Le taux de réponse longitudinal peut être établi sur la base du taux de réponse de chaque interview ou du taux de réponse de chacun des vagues. Pour l'EDTR, nous prévoyons utiliser le taux de réponse des vagues, parce qu'il représente le nombre réel d'enregistrements longitudinaux pondérés du fichier. Cependant, le taux de réponse longitudinal sera ventilé selon que la réponse a été totale ou partielle, afin de pouvoir établir une distinction entre les enregistrements ayant fait l'objet d'une imputation pour une interview donnée et ceux qui n'ont été soumis à aucune imputation.

5. DIFFÉRENCES ENTRE LES RÉPONDANTS ET LES NON-RÉPONDANTS

Le deuxième aspect de la mesure de l'erreur due à la non-réponse est l'analyse des différences entre répondants et non-répondants. Un des avantages d'une enquête longitudinale est que l'on dispose souvent d'une information considérable sur les non-répondants ayant déjà participé aux interviews antérieures. L'EDTR est un supplément à l'Enquête sur la population active, de sorte qu'on possède des renseignements sur les caractéristiques démographiques et l'expérience de travail des personnes n'ayant pas participé à l'interview préliminaire, mais qui ont répondu à l'EPA. Il est donc possible de les différencier des répondants.

Trois types d'études de la qualité des données seront réalisées. Ces études, décrites dans les paragraphes qui suivent, pourront contribuer grandement à une meilleure compréhension de la composition de l'échantillon et peut-être aussi du phénomène d'érosion. Elles pourraient, en outre, aider à orienter le choix de la méthode d'ajustement la plus appropriée pour tenir compte de la non-réponse.

5.1 ANALYSE DES CARACTÉRISTIQUES DES NON-RÉPONDANTS FONDÉE SUR LA RÉPONSE À DES VAGUES PRÉCÉDENTES

Un certain nombre de tableaux ont été produit, qui résument les caractéristiques des non-répondants à l'interview préliminaire et qui nous permettent d'établir des comparaisons avec les répondants. Nous avons observé des différences significatives entre les répondants et les non-répondants selon la province, la région urbaine ou rurale, le groupe d'âge, le niveau d'instruction, le secteur d'activité, la profession, la durée de l'emploi, le travail à plein temps par rapport au travail à temps partiel, le genre de logement, le nombre d'adultes dans le logement et le genre de famille. Un deuxième rapport similaire est en cours pour étudier la

non-réponse aux interviews de la première vague. Les résultats seront présentés dans un rapport spécial sur la qualité des données, et serviront de base pour ajustement de la non réponse de l'échantillon longitudinal.

5.2 COUPLAGE STATISTIQUE DES DONNÉES DE L'ÉCHANTILLON AVEC LES DONNÉES FISCALES

Nous procédons actuellement à un microcouplage des données de l'échantillon du premier panel avec des données du fichier d'impôt pour mesurer la qualité des données du revenu. Ce couplage nous aidera à évaluer certaines des caractéristiques du revenu des non-répondants, en particulier en ce qui concerne leur niveau de revenu et leurs sources de revenu.

Le projet de couplage a été testé avec les données de un essai sur le terrain de l'EDTR effectué en 1993. Les données recueillies lors du test avaient en effet été couplées à celles du fichier fiscal. Certains résultats sont présentés dans les tableaux 3 et 4. Le couplage statistique est fait en fonction du nom de famille, du prénom, du code postal, du sexe et de la date de naissance. Ces renseignements étaient disponibles pour la majeure partie des personnes de l'échantillon (sauf pour quelques non-répondants qui n'ont pas fait l'objet du couplage et qui sont par conséquent exclus des résultats des tableaux).

Le tableau 3 porte sur la déclaration des salaires et traitements. On peut voir que l'échantillon non couplé de l'EDTR diffère grandement de celui qui a fait l'objet d'un couplage. Cependant, les différences entre les répondants et les non-répondants ne sont pas si grandes tant en ce qui concerne le nombre de personnes ayant déclaré des salaires et traitements que le montant moyen déclaré. Par contre, le tableau 4 décrit une toute autre situation.

Tableau 3 Comparaison des salaires et traitements des répondants et des non-répondants de l'échantillon de l'essai 3 réalisé en Ontario

Salaires et traitements	Non-répondants de l'EDTR	Couplage EDTR et fichier fiscal		Données non couplées de l'EDTR
		FICHIER FISCAL	FICHIER FISCAL EDTR	
Nbre de personnes	650	848	848	188
Pourcentage de déclaration	71,4 %	72,5 %	70 %	57 %
Montant moyen déclaré	28 000 \$	27 000 \$	28 000 \$	21 000 \$

Le tableau 4 fait ressortir, chez les répondants, une sous-estimation du pourcentage de gens qui rapportent être bénéficiaires, de même que du montant de la déclaration. Les non-répondants diffèrent là aussi des répondants, la proportion des bénéficiaires de l'assurance-chômage étant plus rapportée sur les données fiscales.

L'assurance-chômage est typiquement sous-estimée dans les enquêtes sur le revenu. Ces résultats peuvent servir à faire la part entre la sous-estimation due à la non-réponse totale et celle qui est due à la non-réponse aux questions. Les microcomparaisons seront aussi utilisées pour estimer l'erreur de réponse, lorsque les définitions des sources de revenu utilisées pour l'enquête et celles qui sont utilisées à des fins fiscales sont compatibles.

Tableau 4 Comparaison du revenu tiré de l'assurance-chômage

Prestations d'assurance-chômage	Non-répondants de l'EDTR	Couplage EDTR et fichier fiscal		Données non couplées de l'EDTR
		FICHER FISCAL	FICHER FISCAL EDTR	
Nbre de personnes	650	848	848	188
Pourcentage de déclaration	18,6 %	15,7 %	12,6 %	13,3 %
Montant moyen déclaré	523 \$	593 \$	498 \$	132 \$

5.3 SUIVI D'UN ÉCHANTILLON DE NON-RÉPONDANTS DE L'INTERVIEW PRÉLIMINAIRE

Un petit sous-échantillon constitué de 12 % des personnes n'ayant pas répondu à l'interview préliminaire de janvier 1993 a été inclus dans l'échantillon de la première vague aux fins d'évaluation de la qualité des données. Le taux de réponse de vague de 1994 pour les personnes comprises dans ce sous-échantillon a été de 66 % (c'est-à-dire la réponse à au moins une des deux interviews sur le travail et sur le revenu). Leurs données seront examinées afin qu'on puisse relever d'éventuelles différences avec les répondants.

6. CONCLUSIONS

Un seul type de taux de réponse ne donne pas une idée précise de la qualité des données dans le cas d'enquêtes d'une grande complexité. D'autres mesures sont nécessaires. Il y a toutefois des désavantages à calculer et à présenter trop de

chiffres, les utilisateurs trouvant plus facile de n'avoir qu'un seul indicateur de réponse (ou quelques-uns, à la rigueur). La réponse à l'EDTR sera bien sûr un sujet très documenté, mais un certain nombre de mesures uniformes devront être appliquées, en particulier le taux de réponse de vague et le taux de réponse longitudinal.

L'équipe de l'EDTR préconise l'adoption d'une définition plus large de «ménage répondant»; par conséquent, la non-réponse aux questions sera plus élevée qu'elle ne le serait si l'on décidait d'utiliser une définition plus restreinte. On voit donc toute l'importance du recours à l'imputation pour tenir compte de la non-réponse aux questions. Nous essaierons de rajuster les calculs de la variance en fonction de l'utilisation ou non de valeurs imputées, sans quoi la variance sera sous-estimée. Des indicateurs seront utilisés pour signaler les cas d'imputation aux utilisateurs.

Il faudrait pousser plus loin l'évaluation des différences entre les répondants et les non-répondants. Aucun indicateur de réponse n'a encore été élaboré pour aider à évaluer les erreurs de mesure.

L'IAO améliore la qualité des données, mais peut aussi le faire de façon «artificielle». Par exemple, dans l'IAO, un enchaînement incorrect (du point de vue des instructions «passez à») est impossible dans la mesure où l'instrument a été soumis à des essais complets. Toutefois, le cheminement choisi peut ne pas refléter la réalité. Ce genre d'erreur peut se révéler beaucoup plus difficile à déceler et à corriger dans une interview assistée par ordinateur que dans un questionnaire rempli à la main. Nous devons penser à des méthodes permettant d'évaluer l'erreur de réponse. D'autres études devront aussi être réalisées pour nous aider à comprendre le mécanisme général de réponse ainsi que les différents éléments qui le composent.

Certaines des mesures dont il a été question dans ce document tiennent compte des particularités de l'EDTR. Il reste que toutes les enquêtes longitudinales ont leurs particularités et qu'il faudra consacrer beaucoup d'efforts pour concevoir des mesures appropriées. Par conséquent, il importe avant tout d'établir des normes de mesure pouvant être adaptées aux caractéristiques propres des diverses enquêtes longitudinales.

RÉFÉRENCES

- [1] Groves, R. (1989), *Survey Errors and Survey Costs*, Wiley Series.
- [2] Comité des méthodes et des normes (1993), *Normes et lignes directrices pour la déclaration des taux de non-réponse*.

ANNEXE 1 DÉFINITION DES UNITÉS AUX FINS DE CALCUL DES TAUX DE RÉPONSE

Définitions adoptées conformément aux normes de Statistique Canada relatives à la déclaration des taux de non-réponse :

Nombre total d'unités	(1)	ensemble des unités comprises dans l'échantillon
Cas résolus	(2)	unités dont on peut établir l'état à la fin de la période de collecte des données
Cas non résolus	(3)	unités dont on ne peut établir l'état à la fin de la période de collecte des données
Unités dans le champ de l'enquête	(4)	unités faisant partie de la population cible de l'enquête
Unités répondantes	(5)	unités ayant fourni des renseignements utilisables
Réponses après refus	(6)	unités qui refusaient de répondre (dans des périodes de collecte courantes ou antérieures) et qui se sont ravisées grâce aux démarches entreprises pour convertir leur refus
Autres répondants	(7)	à l'exclusion des réponses après refus
Unités non répondantes	(8)	unités dans le champ de l'enquête n'ayant pas répondu ou n'ayant pas fourni de renseignements utilisables
Refus	(9)	unités ayant refusé de participer à l'enquête

Unités non contactées	(10)	unités dans le champ de l'enquête qu'il est impossible de contacter
Unités de non-réponse résiduelles	(11)	unités qui n'ont pas répondu en raison de circonstances particulières
Unités hors du champ de l'enquête	(12)	unités qui ne font pas partie de la population cible
Unités inexistantes	(13)	unités hors du champ de l'enquête dont on a déterminé l'inexistence. Dans le cas d'une IAO, il peut s'agir d'unités en double en raison de problèmes de transmission.
Unités provisoirement hors du champ de l'enquête	(14)	unités qui étaient hors du champ de l'enquête au moment de la collecte, mais qui pourraient s'y retrouver ultérieurement (par ex. les personnes en poste à l'étranger pour un an)
Unités hors du champ de l'enquête en permanence	(15)	unités exclues en permanence du champ de l'enquête après la sélection de l'échantillon (par ex. décès)