

N° de catalogue 94-15

**QUALITÉ DES DONNÉES SUR LE REVENU
RECUEILLIES SELON LA MÉTHODE DE L'INTERVIEW
ASSISTÉE PAR ORDINATEUR POUR L'EDTR**

Août 1994

Chantal Grondin, Division des méthodes d'enquêtes sociales

Sylvie Michaud, 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

Cet article a été présenté à la réunion annuelle de l'American Statistical Association, qui s'est tenue à Toronto, en août 1994.

L'interview assistée par ordinateur (IAO) est la méthode de l'avenir. Relativement nouvelle, cette technologie est actuellement utilisée à Statistique Canada pour l'Enquête sur la dynamique du travail et du revenu (EDTR).

L'EDTR est une enquête longitudinale qui a débuté en janvier 1994. Elle consiste en deux interviews par année : l'une en janvier, pour recueillir des données sur le travail, l'autre en mai, pour recueillir des données sur le revenu. Un des buts principaux de l'enquête est de mesurer l'incidence, sur l'activité et le revenu, des changements dans la composition de la famille.

Les questions de la composante «travail» de l'enquête sont passablement linéaires. La programmation de l'application n'a donc posé aucune difficulté. Dans le cas de la composante «revenu», toutefois, la façon de recueillir l'information et la façon dont l'application devait être programmée n'étaient pas évidentes au départ. Par ailleurs, la possibilité de faire des vérifications interactives avec la méthode de l'IAO, vérifications qui sont impossibles à réaliser lors d'interviews «papier et crayon», nous intéressait beaucoup.

Dans cette étude, nous décrivons d'abord l'EDTR. Nous expliquons ensuite comment l'application relative au revenu a été programmée, puis nous exposons certaines des différences que présente l'utilisation de la méthode de l'IAO par rapport à l'interview «papier et crayon», et ce, pour la composante «revenu» de l'enquête. Nous examinons brièvement les procédures d'évaluation utilisées. Il est notamment question du couplage des données recueillies auprès de l'échantillon

d'essai aux données fiscales, des motifs de ce couplage, de la méthode utilisée, du taux d'appariement obtenu et des principales différences entre les appariements et les non-appariements. Nous présentons ensuite les résultats de comparaisons effectuées à un micro-niveau entre les données de l'essai de l'EDTR (recueillies selon la méthode de l'IAO) et les données fiscales pour la même année, ainsi que les résultats de comparaisons menées en parallèle, également à un micro-niveau, entre les données d'une enquête réalisée selon la méthode de l'interview «papier et crayon» et les données fiscales. Nous terminons en formulant certaines recommandations.

TABLE DES MATIÈRES

	Page
1. DESCRIPTION DE L'EDTR	1
2. DIFFÉRENCES ENTRE L'IAO ET L'INTERVIEW «PAPIER ET CRAYON» POUR LA COLLECTE DES DONNÉES SUR LE REVENU ²	
3. ÉVALUATION DE LA QUALITÉ DES DONNÉES : COUPLAGE AVEC LES DONNÉES FISCALES	5
4. COMPARAISONS À UN MICRO-NIVEAU	8
5. RECOMMANDATIONS	18

1. DESCRIPTION DE L'EDTR

L'EDTR est conçue de façon à ce que les membres de l'échantillon longitudinal soient interviewés sur une période de six ans. Des procédures ont été mises en oeuvre pour dépister et interviewer à leur nouvelle adresse (au Canada ou aux États-Unis) les répondants qui déménageraient au cours de ces six ans.

Comme nous l'avons mentionné précédemment, les répondants sont interviewés deux fois l'an. La première interview, menée en janvier de chaque année, sert à recueillir des renseignements détaillés sur l'activité des répondants sur le marché du travail au cours de l'année civile précédente (année de référence). La deuxième interview, qui a lieu en mai de chaque année, permet de recueillir des renseignements sur le revenu provenant de diverses sources qu'ont touché les répondants au cours de l'année de référence. L'interview de mai est en quelque sorte la suite de celle de janvier, puisqu'elle vise la même période de référence. On recueille les données sur le revenu en mai, parce qu'au Canada, c'est vers la fin d'avril, chaque année, que les déclarations de revenus doivent être produites. Il nous a semblé que nous obtiendrions des données de meilleure qualité si l'interview était réalisée à ce moment-là. En fait, nous envoyons par la poste une version imprimée du questionnaire aux répondants avant l'interview et nous encourageons ces derniers à consulter leurs documents fiscaux et à remplir le questionnaire à l'avance pour réduire la durée de l'interview.

En 1993, nous avons fait un essai au cours duquel nous avons simulé tout le processus de collecte des données de l'EDTR au moyen de la méthode de l'IAO. Un échantillon d'environ 1 500 ménages a été prélevé dans deux provinces, soit à Terre-Neuve et en Ontario. Selon le plan de l'EDTR, l'échantillon est prélevé un an avant le premier cycle d'interviews (au début de la première année de référence) dans l'échantillon d'une autre enquête, l'Enquête sur la population active (EPA). Une interview préliminaire est réalisée pour l'EDTR à titre de supplément à l'EPA.

Les renseignements recueillis à ce moment-là sont retransmis aux répondants un an plus tard, lors de l'interview sur le travail, afin de réduire les erreurs de rappel.

Les répondants qui ont participé à l'essai de l'EDTR avaient aussi participé à l'Enquête sur les finances des consommateurs (EFC), également réalisée à titre de supplément à l'EPA. L'EFC est effectuée selon la méthode de l'interview «papier et crayon», et son contenu est très semblable à celui de l'EDTR.

Un des objectifs de l'essai de l'EDTR était de déterminer la meilleure façon de recueillir des données sur le revenu lorsqu'on utilise la méthode de l'IAO.

2. DIFFÉRENCES ENTRE L'IAO ET L'INTERVIEW «PAPIER ET CRAYON» POUR LA COLLECTE DE DONNÉES SUR LE REVENU

Avec la méthode de l'IAO, il est possible d'adapter le questionnaire aux diverses catégories de répondants. Pour l'essai, l'équipe de l'EDTR avait conçu trois questionnaires différents sur le revenu (nous parlons également de «cheminements» ou d'«approches»). Avant l'essai, nous avons envoyé à chaque répondant un carnet sur lequel figuraient toutes les questions de l'interview, des informations utiles sur les éléments de revenu dont il fallait tenir compte et des renvois à des documents fiscaux, le cas échéant. Au moment de l'interview, l'intervieweur demandait au répondant s'il avait rempli le carnet. Si c'était le cas, l'interview était beaucoup plus courte. Le répondant n'avait qu'à indiquer à l'intervieweur les lignes du carnet où il avait inscrit un montant. L'intervieweur pouvait faire défiler le questionnaire à l'écran et inscrire sur les lignes correspondantes les montants que lui indiquait le répondant. Une courte description du poste était donnée à côté de chaque ligne.

Aux répondants qui n'avaient pas rempli le carnet, l'intervieweur demandait s'ils avaient leur déclaration de revenus à portée de la main et, si oui, laquelle. Au Canada, il existe quatre versions de la déclaration de revenus : une déclaration générale, dont tout le monde peut se servir, une déclaration abrégée, une déclaration pour les personnes de 65 ans et plus et une déclaration spéciale. Les trois dernières sont des versions simplifiées de la déclaration générale et s'adressent à certaines catégories de personnes. Idéalement, un poste qui figure sur les quatre versions de la déclaration de revenus porterait le même numéro de ligne sur chaque déclaration. Dans les faits, cependant, il n'en est pas ainsi. Par conséquent, quatre versions du questionnaire ont été programmées de même que quatre séries de renvois aux numéros de lignes des déclarations de revenus. Selon cette approche, les revenus du même type étaient regroupés en blocs, et on posait d'abord au répondant une question générale pour déterminer si l'une des sources de revenus du bloc s'appliquait. Par exemple, un des blocs se rapportait au revenu provenant d'un travail autonome. Si le répondant disait avoir touché un revenu provenant d'un travail autonome pendant l'année de référence, on lui posait ensuite des questions plus précises sur la nature de ce revenu (revenu d'entreprise, de profession libérale, de commissions, d'agriculture, de pêche ou autre). Si le répondant n'avait pas touché un revenu provenant d'un travail autonome, l'intervieweur sautait toutes les questions de ce bloc. Selon la déclaration de revenus qu'utilisait le répondant, certains blocs étaient automatiquement sautés.

Dans le cas des répondants qui n'avaient pas leur déclaration de revenus sous la main, on suivait l'approche «bloc». Cette dernière est semblable à l'approche «déclaration de revenus», que nous venons de décrire, sauf qu'aucun bloc n'est sauté et qu'on ne renvoie pas le répondant à la déclaration de revenus. Ainsi, selon le nombre de réponses affirmatives aux questions générales des blocs, l'interview pouvait être plus longue qu'avec l'approche «carnet» ou l'approche «déclaration de revenus».

Une autre différence importante entre l'IAO et l'interview «papier et crayon» est que l'utilisation d'un ordinateur pour l'interview permet d'effectuer des vérifications interactives qui seraient autrement impossibles à réaliser. Par exemple, chaque fois qu'un montant est entré dans l'ordinateur, une simple vérification du montant en fonction de la fourchette établie peut être faite immédiatement, tandis que le répondant est encore en communication avec l'intervieweur, plutôt qu'à l'étape de la saisie, où il n'est plus possible de vérifier la validité du montant. De plus, avec la méthode de l'IAO, on peut programmer une vérification du revenu total de façon à déterminer si la somme des montants déclarés correspond au revenu total déclaré.

L'IAO et l'interview «papier et crayon» diffèrent également du fait que la première, contrairement à la seconde, permet (en particulier dans le cas d'enquêtes longitudinales) de reporter des renseignements obtenus lors d'une interview antérieure. Par exemple, à partir des renseignements recueillis lors de l'interview sur le travail, en janvier, l'EDTR prévoit l'utilisation de quatre drapeaux correspondant à quatre catégories de revenus : salaires et traitements, prestations d'assurance-chômage, prestations d'aide sociale et indemnités pour accident du travail. Ainsi, le drapeau relatif aux salaires et aux traitements est mis en fonction s'il est établi, en janvier, que le répondant a été un travailleur rémunéré pendant l'année de référence. Au moment de l'interview sur le revenu, une demande de vérification apparaît à l'écran si le répondant ne déclare pas de montant à la question sur les salaires et les traitements. L'intervieweur demande alors au répondant de lui indiquer un montant.

Dans une enquête longitudinale, les renseignements relatifs à l'année précédente peuvent servir à améliorer les vérifications. Par exemple, on peut faire une meilleure vérification du revenu total en se fondant sur le revenu total de l'année précédente.

3. ÉVALUATION DE LA QUALITÉ DES DONNÉES : COUPLAGE AVEC LES DONNÉES FISCALES

Chaque année, les Canadiens produisent leur déclaration de revenus vers la fin d'avril. Les données de la composante «revenu» de l'EDTR sont recueillies au début de mai, période de l'année où les répondants ont davantage en mémoire tous les renseignements qui concernent leur revenu de l'année civile précédente. De plus, selon les résultats d'une enquête antérieure, il est beaucoup plus difficile pour les travailleurs autonomes de répondre à des questions sur le revenu à d'autres moments de l'année, ce qui donne lieu à un taux de non-réponse plus élevé que lorsque l'interview est menée à peu près à la date où les gens doivent produire leur déclaration de revenus.

Bien que les questionnaires d'enquêtes sur le revenu des ménages et la déclaration de revenus présentent des différences sur le plan du contenu, plusieurs évaluations de la qualité des données d'enquêtes sur le revenu comparent les données d'enquêtes avec les données fiscales. Pour l'évaluation de la qualité de nos données, nous avons jugé qu'un couplage avec les données fiscales conviendrait, puisqu'il permettrait une évaluation directe des données recueillies, du moins pour ce qui concerne les sources de revenus comparables à celles qui figurent sur la déclaration de revenus.

Il est impossible de réaliser un couplage direct avec les données fiscales sans le numéro d'assurance sociale (NAS) de chaque répondant. Comme nous n'avions pas ce numéro, nous avons utilisé d'autres renseignements, tels que les nom et prénom du répondant, sa date de naissance, son état matrimonial et son code postal. On s'est servi de ces variables pour réaliser un appariement statistique et coupler les données d'enquête avec les données fiscales. Par souci d'efficacité, nous avons d'abord tenté d'obtenir un appariement exact en utilisant le nom de

famille, la première lettre du prénom, le sexe, la date de naissance, l'état matrimonial et le code postal des répondants. Cette démarche nous a permis d'apparier la moitié de notre échantillon. Pour l'autre moitié, nous nous sommes servi du système CANLINK qui procède à l'appariement de façon probabiliste.

Brièvement, un appariement statistique consiste à coupler les enregistrements qui présentent «la plus forte probabilité» d'appartenir à la même personne, selon la comparaison de certaines variables clés.

Pour éviter des non-appariements dus à des erreurs d'orthographe dans le nom de famille, nous avons utilisé le New York State Intelligence and Identification System (NYSIIS). Ce système code le nom d'après sa représentation phonétique. Par conséquent, deux noms de consonance identique auront le même code NYSIIS.

Pour cette méthode d'appariement, nous avons utilisé comme variables le code NYSIIS attribué au nom de famille des répondants, ainsi que leurs sexe, prénom, nom de famille, date de naissance, état matrimonial et code postal.

Après l'appariement statistique, nous avons extrait du fichier fiscal, pour chaque personne appariée, le nom et le NAS (le cas échéant) de son conjoint. Nous avons pu ainsi apparier quelques personnes de plus, ce qui nous a donné, en définitive, un taux d'appariement de 84 %.

Pour ce qui est des personnes non appariées, nous avons constaté, après examen des données sur le revenu tirées de l'EFC de 1991 pour ces personnes, que la moitié d'entre elles avaient déclaré un revenu nul. Il est donc probable que le fichier fiscal de 1991, dont on s'est servi pour l'appariement (celui de 1992 n'étant pas prêt à ce moment-là), ne contenait aucune donnée sur ces personnes.

**Tableau 1. Caractéristiques des appariements et des non-appariements :
données d'enquête et données fiscales**

GROUPE D'ÂGE	Appariements	Non-appariements
15-19 ans	4,5 %	41,3 %
20-24 ans	9,3 %	6,3 %
25-34 ans	24,4 %	17,9 %
35-54 ans	37,3 %	16,9 %
55 à 64 ans	11,4 %	5,9 %
65 ans et plus	13,2 %	11,6 %
SEXE		
Homme	48,9 %	51,9 %
Femme	51,1 %	48,1 %
ÉTAT MATRIMONIAL		
Marié(e)/en union libre	69,5 %	33,6 %
Célibataire	20,0 %	57,6 %
Veuf/veuve	5,3 %	4,9 %
Séparé(e)/divorcé(e)	5,3 %	3,9 %

Nous avons examiné les caractéristiques des personnes non appariées. La plupart d'entre elles avaient entre 15 et 19 ans et étaient célibataires (41 % des non-appariements) ou étaient âgées de 35 à 54 ans et étaient mariées (12 %). Le tableau 1 montre la répartition des appariements et des non-appariements selon le groupe d'âge, le sexe et l'état matrimonial des répondants.

Au total, 62 % des personnes pour lesquelles un appariement n'a pas été possible vivaient dans des ménages où au moins un membre avait été apparié; le reste des

personnes non appariées faisaient partie de ménages où aucun appariement n'avait été réussi. Cette fois encore, les personnes non appariées dans les ménages partiellement appariés avaient entre 15 et 19 ans ou étaient mariées. Par conséquent, l'hypothèse selon laquelle le fichier fiscal ne contenait aucune données sur ces personnes, et que celles-ci étaient plutôt déclarées comme personnes à charge sur la déclaration de revenus de leur conjoint, de leur père ou de leur mère est très plausible, du moins pour une bonne proportion de ces personnes.

4. COMPARAISONS À UN MICRO-NIVEAU

Notre évaluation de la qualité des données nous a permis de comparer, pour un échantillon de personnes et à un niveau élevé de détail, d'une part, des données d'enquête recueillies selon la méthode de l'interview «papier et crayon» avec les données fiscales pour 1991 et, d'autre part, des données d'enquête recueillies selon la méthode de l'IAO avec les données fiscales pour 1992. Dans l'analyse qui suit, nous nous concentrons sur les comparaisons effectuées à un micro-niveau pour montrer comment les réponses obtenues au moyen de l'IAO diffèrent de celles qu'on recueille selon la méthode de l'interview «papier et crayon».

Les comparaisons présentées ici sont limitées à trois variables : salaires et traitements (ST), prestations d'assurance-chômage (AC) et intérêts et dividendes (ID). Le choix de ces variables a été motivé par le fait que l'IAO permet une vérification interactive des variables ST et AC, contrairement à l'interview «papier et crayon». Comme nous l'avons mentionné précédemment, selon les réponses obtenues lors de notre interview de janvier sur le travail, un drapeau était mis en fonction si un répondant avait été un travailleur rémunéré pendant l'année de référence et un autre drapeau était mis en fonction si un répondant avait été en chômage cette année-là. Une demande de vérification s'affichait alors à l'écran de l'intervieweur lors de l'interview sur le revenu dans les cas où le répondant ne

déclarait aucun montant relativement aux ST ou à l'AC alors qu'il devait le faire. L'intervieweur devait donc demander au répondant de lui préciser un montant. Nous espérons que cette vérification allait résoudre certains problèmes de sous-déclaration, en particulier dans le cas des prestations d'AC, qu'on soupçonne d'être sous-déclarées dans l'EFC. Quant au choix de la variable ID pour notre comparaison, il s'explique par le fait que, même s'il n'existe pour cette variable aucune vérification semblable à celle des ST ou de l'AC, cette catégorie de revenu pose également des problèmes de sous-déclaration.

Avant d'examiner les résultats des comparaisons à un micro-niveau, il importe de mettre en évidence les limites de cette étude. Premièrement, les répondants de l'essai de l'EDTR avaient fait partie de l'échantillon de l'EFC l'année précédente. Il se peut donc qu'il y ait un effet de conditionnement qu'il nous est impossible de mesurer. Deuxièmement, les questionnaires de l'EDTR et de l'EFC étaient assez semblables en ce qui concerne le revenu, mais celui de l'EDTR contenait des questions plus détaillées pour certaines sources de revenu; en outre, l'interview comportait des questions sur le patrimoine. Il est possible que cela ait alourdi le fardeau des répondants. Troisièmement, la documentation envoyée aux répondants avant l'interview n'était pas la même pour les deux enquêtes. La documentation de l'EDTR était plus facile à utiliser et on faisait davantage référence à la déclaration de revenus. Quatrièmement, le taux de réponse à l'enquête réalisée selon la méthode de l'IAO était d'environ 67 %, tandis que le taux correspondant pour l'enquête réalisée au moyen d'une interview «papier et crayon» était d'environ 83 %. D'après notre propre expérience avec l'utilisation de la méthode de l'IAO, nous avons constaté que cette dernière introduit un nouveau facteur de non-réponse : l'échec des transmissions (nous sommes conscients qu'il s'agit là d'un problème qui peut ne survenir que lors d'enquêtes décentralisées réalisées selon la méthode de l'IAO). Comme les cas sont transmis entre

ordinateurs, il est possible que certains se perdent ou que leur réception soit retardée, de sorte qu'ils sont considérés comme des cas de non-réponse.

Malgré ces limites, il nous a semblé que l'évaluation pouvait quand même donner une idée de la qualité des données recueillies selon la méthode de l'IAO.

Les comparaisons à un micro-niveau ont produit des résultats intéressants. Le tableau 2 montre, pour chaque variable à l'étude, le pourcentage de personnes dont le revenu est nul selon les données d'enquête et selon le fichier fiscal, le pourcentage de celles dont le revenu est le même selon les données d'enquête et selon le fichier fiscal et le pourcentage de celles dont le revenu selon les données d'enquête diffère de celui qui figure dans le fichier fiscal.

Le pourcentage de personnes dont le revenu provenant des salaires et traitements était nul selon les données d'enquête et les données fiscales était légèrement plus élevé dans l'échantillon de l'IAO. Cette proportion supérieure tient davantage à ce que les années d'enquête sont différentes qu'à la méthode de collecte des données elle-même. De fait, on peut observer la même tendance quand on recoupe dans un tableau des données fiscales de 1991 et de 1992 : le pourcentage de personnes qui n'ont déclaré aucun salaire ou traitement en 1992 est supérieur de 5 % au pourcentage correspondant en 1991. L'écart est probablement attribuable à un facteur économique, le taux de chômage ayant été plus élevé en 1992 qu'en 1991. On peut faire la même observation pour ce qui concerne les intérêts et les dividendes, quoique les écarts entre les deux années soient plus grands. Par contre, en ce qui a trait à l'assurance-chômage, les résultats pour les deux années sont très semblables; ils sont aussi à peu près semblables quand on ne considère que les données fiscales.

Tableau 2. Comparaisons à un micro-niveau entre les données d'enquête et les données fiscales, selon la méthode de l'interview «papier et crayon» et la méthode de l'IAO

ST	Papier et crayon	IAO
données d'enquête et données fiscales = 0	24,7 %	29,6 %
données d'enquête = données fiscales	32,7 %	32,9 %
données d'enquête différentes des données fiscales	42,5 %	37,5 %
AC		
données d'enquête et données fiscales = 0	69,1 %	70,5 %
données d'enquête = données fiscales	13,7 %	11,9 %
données d'enquête différentes des données fiscales	17,2 %	17,6 %
ID		
données d'enquête et données fiscales = 0	50,7 %	58,8 %
données d'enquête = données fiscales	16,2 %	17,9 %
données d'enquête différentes des données fiscales	33,0 %	23,2 %

En général, les données d'enquête semblent plus près des données fiscales pour les variables ST et ID avec la méthode de l'IAO qu'avec la méthode «papier et crayon». On peut le constater par le plus faible pourcentage de personnes dans la catégorie «données d'enquête différentes des données fiscales» avec la méthode de l'IAO qu'avec la méthode «papier et crayon». Les résultats sont à peu près les mêmes pour la variable AC avec les deux méthodes de collecte. Dans le tableau 3 ci-après, on examine le pourcentage de personnes pour lesquelles on trouve des montants non nuls dans les données d'enquête ou dans les données fiscales.

Tableau 3. Montants non nuls dans les données d'enquête ou dans les données fiscales

ST	Papier et crayon	IAO
Montants identiques à 5 % près	61,9 %	67,6 %
Écart de plus de 5 %	21,8 %	24,0 %
Montant dans données fiscales seulement	10,4 %	3,4 %
Montant = NSP dans l'enquête	3,8 %	3,7 %
Montant dans données d'enquête seulement	2,1 %	1,4 %
AC		
Montants identiques à 5 % près	51,3 %	50,3 %
Écart de plus de 5 %	21,8 %	29,5 %
Montant dans données fiscales seulement	16,0 %	10,7 %
Montant = NSP dans l'enquête	8,3 %	8,1 %
Montant dans données d'enquête seulement	2,6 %	1,3 %
ID		
Montants identiques à 5 % près	36,4 %	47,4 %
Écart de plus de 5 %	18,1 %	20,2 %
Montant dans données fiscales seulement	41,4 %	21,2 %
Montant = NSP dans l'enquête	1,8 %	5,8 %
Montant dans données d'enquête seulement	2,4 %	5,5 %

Il semble y avoir une plus grande concordance entre les données d'enquête et les données fiscales avec la méthode de l'IAO qu'avec la méthode «papier et crayon», sauf pour la variable AC, où les écarts ne sont pas significatifs.

En outre, la sous-déclaration semble moins forte avec la méthode de l'IAO qu'avec la méthode «papier et crayon» pour les trois sources de revenu. Il est possible que les drapeaux pour les variables ST et AC aient aidé le répondant à se souvenir des différents types de revenus qu'il avait touchés pendant l'année. Pour ce qui est de la variable ID, les questions sur le patrimoine dans le carnet, en particulier celles qui portaient sur les comptes bancaires et les économies, peuvent avoir porté le répondant à considérer davantage toutes ses sources de revenus, ce qui a ainsi donné lieu à une moins forte sous-déclaration des revenus de cette source.

Le pourcentage de personnes n'ayant pas déclaré de montant parce qu'elles ne le connaissaient pas ou qu'elles refusaient de le faire est plus élevé avec la méthode de l'IAO qu'avec la méthode «papier et crayon» pour la variable ID, mais il est presque le même avec les deux méthodes pour les autres variables. Pour la variable ID, la proportion de personnes pour lesquelles un montant figure dans les données d'enquête seulement est plus forte avec la méthode de l'IAO.

Pour l'EDTR, nous avons croisé les mêmes résultats avec la variable «cheminement» pour déterminer si la qualité des données était différente selon l'approche suivie par le répondant. Le tableau 4 ci-après montre les résultats de la comparaison.

Tableau 4. Comparaison des données fiscales et des données d'enquête recueillies au moyen de l'IAO, selon l'approche suivie, pour les montants non nuls

ST	Carnet		Décl. de rev.	BLOC
Montants identiques à 5 % près	83,2 %		85,8 %	36,2 %
Écart de plus de 5 %	11,8 %	11,8 %		47,2 %
Montant dans données fiscales seulement	3,4 %	1,2 %	4,9 %	
Montant = NSP dans l'enquête	0,0 %	0,0 %	10,6 %	
Montant dans données d'enquête seulement	1,7 %	1,2 %	1,2 %	
AC				
Montants identiques à 5 % près	73,7 %		80,4 %	16,41 %
Écart de plus de 5 %	15,8 %	10,7 %		50,0 %
Montant dans données fiscales seulement	7,9 %	7,1 %	14,8 %	
Montant = NSP dans l'enquête	0,0 %	1,8 %	18,0 %	
Montant dans données d'enquête seulement	2,6 %	0,0 %	0,8 %	

ID

Montants identiques à 5 % près	66,3 %	70,3 %	9,5 %
Écart de plus de 5 %	16,0 %	17,6 %	29,3 %
Montant dans données fiscales seulement	13,1 %	8,8 %	36,7 %
Montant = NSP dans l'enquête	0,0 %	1,1 %	15,7 %
Montant dans données d'enquête seulement	4,6 %	2,2 %	8,8 %

Il apparaît évident que les approches «carnet» et «déclaration de revenus» donnent de meilleurs résultats que l'approche «bloc». Il est probable que les personnes qui prennent volontiers le temps de remplir leur carnet le font en consultant leur déclaration de revenus. Par contre, les personnes qui n'utilisent pas leurs documents fiscaux donnent davantage des montants approximatifs et oublient vraisemblablement certaines sources de revenus. De fait, nous avons également comparé les renseignements obtenus au moyen des trois approches. Selon les résultats de cette comparaison, les personnes qui ont suivi l'approche «carnet» ou l'approche «déclaration de revenus» ont déclaré plus de sources de revenus que les personnes qui ont suivi l'approche «bloc».

Les résultats indiqués au tableau 4 montrent une très grande concordance entre les données d'enquête et les données fiscales pour l'approche «carnet» et l'approche «déclaration de revenus». De plus, le pourcentage de personnes n'ayant déclaré dans l'enquête aucun montant pour une source donnée, mais pour lesquelles un revenu de cette source figurait dans le fichier fiscal, est beaucoup plus élevé parmi celles qui ont suivi l'approche «bloc».

Le tableau qui suit est similaire au tableau 4, mais il se rapporte aux personnes pour lesquelles un drapeau signalait la présence ou l'absence de revenu pour les variables ST et AC. Par conséquent, nous n'avons retenu pour les fins de la comparaison que les personnes qui ont répondu à l'interview sur le travail, car c'est lors de cet interview que les drapeaux ont été créés.

Tableau 5. Répartition des montants selon la présence ou l'absence d'un drapeau

ST	Présence	Absence
Montants identiques à 5 % près	68,3 %	7,9 %
Écart de plus de 5 %	24,4 %	1,2 %
Montant dans données fiscales seulement	1,3 %	4,3 %
Montant = NSP dans l'enquête	3,7 %	0,6 %
Montant dans données d'enquête seulement	0,5 %	2,1 %
Données d'enquête et données fiscales = 0	1,9 %	83,8 %
AC		
Montants identiques à 5 % près	53,9 %	5,6 %
Écart de plus de 5 %	31,4 %	2,7 %
Montant dans données fiscales seulement	3,7 %	2,9 %
Montant = NSP dans l'enquête	7,9 %	1,2 %
Montant dans données d'enquête seulement	0,0 %	0,5 %
Données d'enquête et données fiscales = 0	3,1 %	87,1 %

Ce tableau montre des résultats très intéressants. D'abord, dans la plupart des cas où il y avait un drapeau (indiquant qu'un montant devait être fourni), un montant avait en général été déclaré lors de l'interview sur le revenu (93,2 % pour la variable ST et 85,3 % pour la variable AC, chiffres correspondant à la somme des

pourcentages qui se rapportent aux première, deuxième et cinquième catégories du tableau). De plus, lorsqu'un montant avait été déclaré dans l'enquête, les données fiscales confirment le bien-fondé du drapeau dans presque tous les cas pour la variable ST et dans tous les cas pour la variable AC. En outre, dans les cas où il y avait un drapeau mais qu'aucun montant n'avait été déclaré, les données fiscales indiquent, dans la majorité des cas, qu'il aurait dû y avoir un montant (dans les cas où le répondant ne savait pas le montant ou a refusé de le déclarer dans l'enquête, les données fiscales indiquent un montant, sauf dans un cas pour la variable AC.). Dans quelques cas seulement, le drapeau paraît avoir été en erreur. Par conséquent, il semble que la présence du drapeau soit en général un bon prédicteur de la présence d'un montant.

Par ailleurs, quand il n'y avait pas de drapeau, mais qu'un montant avait été déclaré lors de l'interview sur le revenu, les données fiscales confirment qu'un montant devait effectivement l'être dans 81,2 % des cas pour la variable ST et dans 94,3 % des cas pour la variable AC (chiffres correspondant à la somme des pourcentages se rapportant aux première et deuxième catégories du tableau divisée par la somme des pourcentages se rapportant aux première, deuxième et cinquième catégories). Ces résultats indiquent que, lors de l'interview sur le travail, les répondants n'ont pas dit qu'ils avaient été des travailleurs rémunérés (ou des bénéficiaires de l'assurance-chômage), comme ils auraient dû le faire. Par conséquent, il est possible qu'il y ait lieu dans ces cas d'imputer le drapeau ainsi que tous les renseignements qui s'y rapportent dans l'interview sur le travail. Quant aux cas où il n'y avait ni drapeau ni montant déclaré dans l'enquête, le fichier fiscal indique un montant pour un petit pourcentage d'entre eux (4,3 % pour la variable ST et 2,9 % pour la variable AC). Il nous était impossible de savoir qu'un montant aurait dû être déclaré dans ces cas-là.

Nous avons examiné les montants indiqués dans les fichiers fiscaux pour les personnes qui ont refusé de déclarer un montant dans l'enquête. Le fichier fiscal contenait un montant dans tous les cas pour la variable ST et dans tous les cas sauf un pour la variable AC. Nous pouvons donc envisager l'imputation d'un montant dans les cas où les répondants refusent de déclarer une source de revenu.

Nous avons également étudié ces résultats selon l'approche suivie pour voir si les réponses étaient différentes. Nous avons noté que, si le drapeau semble favoriser la déclaration d'un montant indépendamment de l'approche suivie, l'exactitude du montant déclaré varie beaucoup d'une approche à l'autre. Par exemple, parmi les personnes qui ont suivi l'approche «bloc» et pour lesquelles il y avait un drapeau pour la variable ST, seulement 37 % ont déclaré un montant à 5 % près de celui que contient le fichier fiscal. Pour les approches «carnet» et «déclaration de revenus», le pourcentage correspondant s'établit à environ 85 %. On a observé le même effet dans le cas des personnes pour lesquelles il y avait un drapeau pour la variable AC.

5. RECOMMANDATIONS

La mise en oeuvre d'enquêtes faisant appel à la méthode de l'IAO est encore relativement récente à Statistique Canada et il reste encore beaucoup d'essais à réaliser avant que nous puissions tirer des conclusions définitives. Il nous a été impossible de tout comparer convenablement la méthode de l'IAO et la méthode «papier et crayon»; nos conclusions sont donc limitées et certains effets peuvent se confondre. Cependant, notre essai nous a permis de prendre certaines décisions qui influent sur le plan qui sera mis en oeuvre pour la réalisation de l'enquête proprement dite.

Il y a bel et bien une différence dans la qualité des données selon la méthode qu'utilisent les personnes pour fournir les renseignements demandés.

Les personnes qui se sont servies du carnet ou de leur déclaration de revenus ont fourni des renseignements de meilleure qualité. Nous ne savons pas de façon certaine, cependant, si le fait d'avoir conçu trois approches différentes pour faciliter la tâche aux répondants et aux intervieweurs a contribué de façon importante à l'amélioration de la qualité des données. Dans le cas des interviews «papier et crayon» pour les enquêtes sur le revenu, on estime habituellement que 40 % des personnes remplissent le questionnaire avant l'interview (ce qui correspond à l'approche «carnet» dans l'EDTR). Par conséquent, la meilleure qualité des données recueillies selon la méthode de l'IAO peut, en partie, être attribuée au fait que, parce qu'on avait conçu une application axée sur la déclaration de revenus, les intervieweurs ont été portés à demander aux répondants de consulter leurs documents fiscaux.

Bien que les données semblent de meilleure qualité quand elles sont recueillies selon l'approche «carnet» ou l'approche «déclaration de revenus», une seule approche, semblable à celle du carnet, sera utilisée pour l'enquête proprement dite. Ce choix s'explique par un certain nombre de facteurs : d'abord, parce que nous voulons simplifier l'instrument de collecte le plus possible pour les intervieweurs, ensuite, parce que les taux de réponse lors de l'essai de la méthode de l'IAO ont été beaucoup plus faibles que ceux que l'on observe habituellement dans les enquêtes sur le revenu réalisées selon la méthode de l'interview «papier et crayon».

Le taux de réponse plus faible est peut-être dû au fait que, pour obtenir des renseignements sur des catégories de revenu semblables à celles des fichiers fiscaux, il soit nécessaire de décomposer ces catégories. Par ailleurs, nous avons aussi tenté de recueillir des données sur le patrimoine lors de l'essai. La

combinaison des deux parties a donné lieu à un nombre de questions deux fois plus élevé pour l'application IAO que pour l'interview «papier et crayon». Nous pensons que le fait de réduire le nombre de questions fera probablement augmenter le taux de réponse. Toutefois, comme l'approche «déclaration de revenus» a produit de bons résultats du point de vue de la qualité des données, on devrait former les intervieweurs à encourager les répondants à consulter leurs documents fiscaux chaque fois que possible.

On a utilisé la technique de l'interview avec rétro-information grâce aux drapeaux mis en fonction dans le système informatique. L'interview réalisée en janvier a servi à recueillir des renseignements détaillés sur le travail; pour cinq sources de revenus, on a pu poser un drapeau pour indiquer si un montant devait être déclaré ou non. Lors de l'interview sur le revenu, si aucun montant n'était déclaré pour l'une ou l'autre de ces sources, le drapeau déclenchait une vérification, selon laquelle l'intervieweur demandait au répondant s'il avait oublié la source en question. Les résultats de l'analyse ayant porté sur les drapeaux a démontré:

- i) Quand un drapeau avait été mis à «vrai» (pour signifier qu'un montant devait être déclaré), un montant avait été déclaré, ou la réponse «Ne sait pas» indiquée, dans 96 % des cas et plus. Cela a permis de réduire de beaucoup la proportion de cas où aucun montant n'avait été déclaré dans l'enquête, mais où un montant figurait dans le fichier fiscal. Quand un drapeau avait été mis à «vrai», mais qu'aucun montant n'avait été déclaré, il semble que, dans la moitié des cas, le drapeau eût été mis par erreur (aucun montant ne figurait dans le fichier fiscal).
- ii) Quand le drapeau avait été mis à «faux» (pour signifier qu'aucun renseignement recueilli en janvier ne nous portait à croire qu'un montant devait être déclaré) et qu'un montant avait été déclaré pour cette source, les

données fiscales semblaient indiquer que nous devions tenir compte du montant (un montant figurait également dans le fichier fiscal pour la même source).

- iii) Même si l'ajout de drapeaux a contribué à améliorer la déclaration des sources de revenu, certains problèmes liés à la qualité des données demeurent pour les répondants interviewés selon l'approche «bloc». Les montants sont souvent arrondis ou la réponse «Ne sait pas» est donnée au lieu du montant.

Par conséquent, l'application dont nous nous servons pour l'enquête proprement dite comprendra des drapeaux pour faciliter l'interview. Quant à la qualité des données, nous avons l'intention de poursuivre notre évaluation des données de l'essai et de faire une évaluation sur une base continue.