

N° 97-06 au catalogue

**ENQUÊTE SUR LA DYNAMIQUE DU TRAVAIL ET DU
REVENU: STRATÉGIE RELATIVE AU TRAITEMENT DES
DONNÉES DE LA VAGUE 1 SUR LE REVENU**

Numéro d'enregistrement du produit 75F0002M

Mars 1997

Élaine Fournier, Division des enquêtes-ménages

Tracey Leesti, Division des enquêtes-ménages

Mylène Lavigne, 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 C6, Statistique Canada, Ottawa (Ontario), Canada, K1A 0T6; par INTERNET: DYNAMIQUE@STATCAN.CA; par téléphone au (613) 951-4633; ou par télécopieur au (613) 951-3012.

SOMMAIRE

Ce document présente la méthode de vérification et d'imputation qui a été retenue dans l'exploitation des données sur le revenu de la première phase de l'Enquête sur la dynamique du travail et du revenu. Il vise à informer les utilisateurs que ces procédures intéressent et à fournir pour l'avenir un constat des raisons pour lesquelles certaines mesures ont été prises. À l'heure même où ce document était préparé, le personnel d'enquête s'en servait comme outil de référence dans l'exploitation des données de la deuxième phase.

TABLE DES MATIÈRES

	Page
1. Mise en contexte	1
2. Questions d'ordre conceptuel	2
2.1 Système d'exploitation	2
2.2 Fusion des données fiscales et des données d'enquête : définition des catégories de revenu et du revenu total	5
2.3 Liens entre le travail et le revenu	10
2.4 Quelques principes fondamentaux de correction et d'imputation	15
3. Correction des données d'enquête	16
3.1 Classement des observations des intervieweurs et réaction	16
3.2 Comparaison du revenu total indiqué à la somme des éléments	17
3.3 Vérifications	19
3.3.1 Prestation fiscale pour enfants	19
3.3.2 Pension de sécurité de la vieillesse	19
3.3.3 Pension de l'employeur	20
3.3.4 Régime de pensions du Canada/Régime de rentes du Québec	21
3.3.5 Cas «à préciser»	29
3.4 Étude des valeurs extrêmes	29
3.5 Attribution de codes-réponses aux données sur le revenu	31
3.6 Compatibilité entre la catégorie de travailleurs et le revenu d'emploi	34

4.	Appariement des fichiers d'impôt	36
4.1	Description sommaire de la méthode utilisée : appariement direct ou appariement statistique	37
4.2	Quelques taux d'appariement et plans d'évaluation	38
5.	Méthodes d'imputation et de calcul	49
5.1	Système généralisé de vérification et d'imputation (SGVI)	49
5.2	Imputation en fonction du plus proche voisin à l'aide du SGVI	50
5.2.1	Imputation : «receveurs — non-réponse partielle»	51
5.2.2	Imputation : «receveurs — non-réponse totale»	53
5.2.3	Imputation déterministe	54
5.2.4	Imputation de l'impôt fédéral et provincial sur le revenu	54
5.3	Évaluation de l'imputation en fonction du plus proche voisin à l'aide du SGVI	55
5.4	Revenu total	60
6.	Revenu familial et faible revenu	61

1. MISE EN CONTEXTE

Du point de vue de l'exploitation, les données sur le revenu de l'EDTR sont inhabituelles, et ce, à deux égards. Premièrement, les répondants se sont vu offrir la possibilité d'autoriser l'accès à leur dossier fiscal au lieu de participer à une interview sur leur revenu. Comme les données de certains répondants viennent des dossiers fiscaux et que celles d'autres répondants ont été recueillies dans le cadre d'interviews, il a fallu fusionner des données obtenues de deux manières différentes pour les exploiter. Les différences de contenu des données fiscales et des données recueillies ne sont pas accidentelles. Les sources de données explicites sur le revenu qui servent dans les interviews sont essentiellement les mêmes que celles de l'Enquête sur les finances des consommateurs. Les décisions relatives au contenu s'appuient sur un raisonnement théorique que tempère la valeur de données recueillies dans une interview téléphonique auprès de répondants qui ont ou non accès à des documents financiers. Par contraste, les dossiers fiscaux contiennent essentiellement les données nécessaires au calcul du revenu imposable, des déductions et des crédits d'impôt. Il a donc fallu prendre une série de décisions avant d'unir ces sources de données sur le revenu.

Les données sur le revenu de l'EDTR présentent en outre un caractère inhabituel en ce qu'elles sont longitudinales. Quand l'exploitation des données de la deuxième phase va commencer, il faudra examiner un certain nombre de questions touchant l'emploi des données des années antérieures dans la vérification ou l'imputation des données de l'année courante.

2. QUESTIONS D'ORDRE CONCEPTUEL

2.1 Système d'exploitation

L'élément «revenu» du système d'exploitation de l'EDTR est divisé en quatre principales parties ou «entités» (figure 1). La première entité, INCCOL, contient des renseignements sur le revenu qui ont été recueillis dans le cadre d'une interview. L'entité contient des données d'enquête brutes qui sont réservées au Bureau, et les variables qu'elle contient ne figurent pas dans le fichier à grande diffusion. Un dossier a été créé pour chacune des personnes auprès desquelles des données sur le revenu ont été recueillies (code-réponse complet ou partiel).

La deuxième entité, INCTAX, contient des données fiscales qui ont été extraites des fichiers d'impôt de Revenu Canada. Comme l'entité INCCOL, l'entité INCTAX est réservée au Bureau, et les variables qu'elle contient ne figurent pas dans le fichier à grande diffusion.

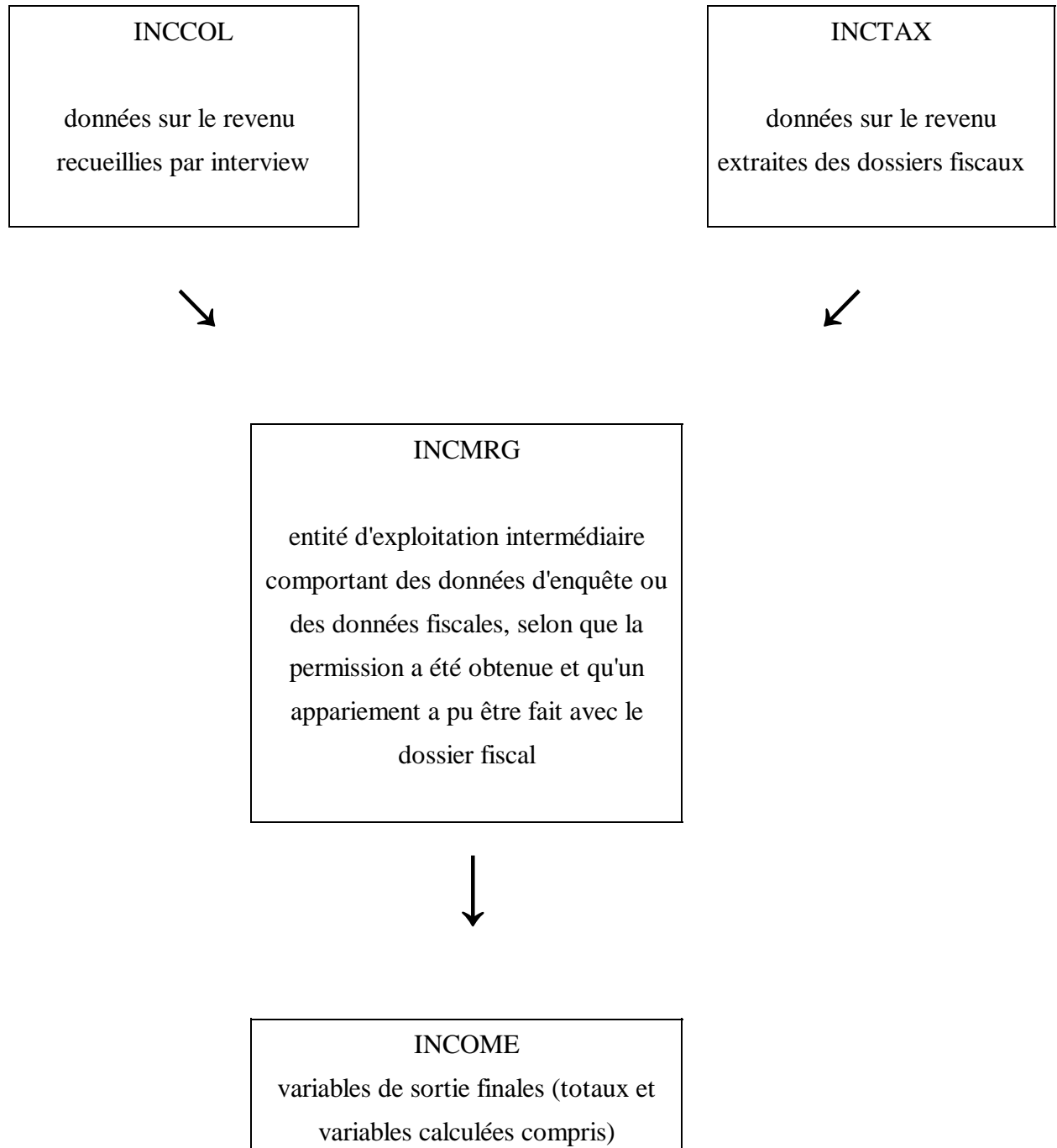
La troisième entité du modèle de données sur le revenu est INCMRG. Cette entité permet de «fusionner» les données d'enquête et les données fiscales en un ensemble unique de variables. Si un répondant a autorisé la consultation de son dossier fiscal et que l'appariement a pu être fait, des renseignements tirés de ses déclarations d'impôt sont gardés dans cette entité. Si une personne n'a pas donné cette autorisation ou si l'appariement n'a pas pu être fait, ce sont les données d'enquête qui sont gardées dans cette entité. L'entité contient un dossier pour chaque personne qui avait 16 ans et plus à la fin de l'année de référence.

L'entité INCOME est la dernière entité utilisée dans la diffusion des données. Comme l'entité INCMRG, elle contient des données fiscales et des données

d'enquête. L'entité comprend aussi des valeurs imputées, divers revenus totaux et des variables calculées.

Les données de 1993 ont toutes été obtenues par interview, puisque, à cette époque, on ne demandait pas encore aux répondants l'autorisation de consulter leur dossier fiscal. Les répondants (code-réponse complet ou partiel) avaient donc tous un dossier dans l'entité INCCOL. C'est en mai 1995 qu'on a demandé pour la première fois aux répondants la permission de consulter leur dossier fiscal. Cette autorisation étant rétroactive, elle s'appliquait aux six années d'existence de l'échantillon permanent. On a donc utilisé les données de 1993 dans le cas des répondants qui avaient donné l'autorisation, et les données d'enquête pour les autres répondants.

Figure 1. Quatre entités de l'élément «revenu» du système d'exploitation de l'EDTR



2.2 Fusion des données fiscales et des données d'enquête : définition des catégories de revenu et du revenu total

Comme on l'a noté ci-dessus, les questions de l'interview sur le revenu et les lignes de la déclaration d'impôt ne correspondent pas parfaitement. Le tableau 1 ci-dessous montre comment on est parvenu à un ensemble unique de catégories de revenu à partir de ces deux sources de données.

Table 1 : Définitions utilisées dans la fusion des données sur le revenu de l'EDTR et des données fiscales de 1993

Variable	Enquête	Données fiscales	Commentaires sur la comparabilité
1. Revenu d'emploi			
Rémunérations	Rémunérations <i>Q1</i>	Revenus d'emploi (rémunérations) <i>L101</i>	Avantages sociaux : en 1993, les avantages sociaux ne figuraient pas dans les données d'enquête (en principe du moins), mais ils entraient dans les données fiscales. Depuis 1994, les avantages sociaux sont ajoutés aux rémunérations dans les deux sources de données.
Revenu net d'un travail agricole indépendant	Revenu net d'un travail agricole indépendant <i>Q2</i>	Revenus d'agriculture <i>L141</i>	Compatibilité vérifiée en fonction de la catégorie de travailleurs et de l'activité économique déclarées dans l'interview de l'enquête. Les personnes déclarant un revenu ici devraient avoir une exploitation agricole non constituée en corporation.
Revenu net d'un travail indépendant non agricole	Revenu net d'un travail indépendant non agricole <i>Q3</i>	Revenus d'entreprise, de profession libérale, de commissions, de pêche <i>L135, L137, L139, L143</i>	Compatibilité vérifiée en fonction de la catégorie de travailleurs et de l'activité économique.

2. Revenu de placements

Revenu de placements (y compris dividendes nets)	Intérêts <i>Q4</i> + dividendes nets <i>Q5</i> + autres revenus de placements <i>Q7</i>	Dividendes (<i>L120</i> (x 0,8)) + intérêts et autres revenus de placements (<i>L121</i>) + revenus nets de société de personnes (<i>L122</i>) + revenus nets de location (<i>L126</i>)	<p>- Les données fusionnées ne comprennent qu'une grande catégorie de revenus de placements. Il y a de nombreuses différences entre les deux sources à l'intérieur même de la catégorie des placements.</p> <p>- Les intérêts tirés de prêts et d'hypothèques et le revenu d'un fonds en fiducie ou d'une succession sont ajoutés aux autres revenus de placements dans l'enquête. Ces éléments ne figurent pas expressément dans la déclaration d'impôt.</p> <p>- Les dividendes nets ont été recueillis dans le cadre de l'enquête. Les dividendes imposables ont été obtenus en multipliant les dividendes nets (c'est-à-dire réels) par 125 %. Depuis 1994, des données sur les dividendes imposables sont recueillies dans le cadre de l'enquête.</p> <p>- La seule différence entre les deux variables des placements tient au traitement des dividendes. Les revenus de placements sont utilisés dans le calcul du revenu monétaire total. Les revenus de placements imposables sont utilisés dans le calcul de l'impôt exigible.</p>
Revenus de placements imposables (dividendes imposables compris)	Intérêts <i>Q4</i> + dividendes imposables <i>Q5</i> (x 125 %) + autres revenus de placements <i>Q7</i>	Dividendes (<i>L120</i>) + intérêts et autres revenus de placements (<i>L121</i>) + revenus nets de société de personnes (<i>L122</i>) + revenus nets de location (<i>L126</i>)	

3. Transferts des gouvernements

Prestation fiscale pour enfants (PFE)	Prestation fiscale pour enfants <i>Q8</i>	Ne se trouve pas sur la T1, mais dans un fichier distinct de Revenu Canada	Dans l'enquête, le montant a été calculé en fonction du nombre d'enfants et du revenu familial net s'il n'a pas été déclaré.
---------------------------------------	--	--	--

Pension de sécurité de la vieillesse, supplément de revenu garanti, allocation de conjoint	PSV/SRG/ AC Q9	PSV (L113) + Suppléments fédéraux (L146)	La PSV est dissociée du SRG et de l'AC, car elle est imposable, alors que le SRG et l'AC ne le sont pas. Dans l'enquête, le montant donné en réponse à la question 9 peut être fractionné, puisque la PSV est généralement un montant fixe.
Régime de pensions du Canada ou Régime de rentes du Québec	RPC/RRQ Q10	RPC/RRQ (L114)	
Prestations d'assurance-chômage	Prestations d'a.-c. Q11	Prestations d'a.-c. (L119)	
Assistance sociale	AS et SRP Q12	Assistance sociale (L145)	
Indemnités pour accidents du travail	IAT Q13	IAT (L144)	
Crédit pour la taxe sur les produits et services	Crédit pour la TPS Q14 - imputé si le répondant est admissible et que la TPS n'a pas été déclarée	Ne figure pas sur la T1. Calculé en fonction du revenu familial.	En 1994, on pourra utiliser le fichier de la TPS au lieu de faire le calcul, comme cela se fait dans le cas des PFE. Le crédit pour la TPS est également calculé pour les répondants admissibles de l'enquête.
Transferts imposables des gouvernements			PSV + RPC/RRQ + A.-C. Ne comprend pas les autres revenus imposables de source gouvernementale, qui sont ajoutés aux autres revenus imposables.

Transferts non imposables des gouvernements			PFE + SRG/AC + crédits pour la TPS + IAT.
---	--	--	---

4. Revenu de pensions

Pension de l'employeur, retraits de REER et rente d'un FERR	Pensions d'employeurs <i>Q18</i> + Rentes de REER et retraits de FERR <i>Q19</i>	Autres pensions et pensions de retraite (<i>L115</i>) + Revenu d'un REER (rente ou retrait) pour les personnes de 65 ans et plus (<i>L129</i>)	Dans les données fiscales, il ne peut pas être fait de distinction entre les rentes de REER et les retraits de REER non échus (qui ne sont pas assimilés à un revenu monétaire). Il a été postulé que le montant indiqué à la L129 est une rente dans le cas des personnes de 65 ans et plus et un retrait dans le cas des personnes qui n'ont pas atteint cet âge.
---	---	---	---

5. Autres revenus imposables

Pension alimentaire et soutien des enfants	Pension alimentaire, allocation de séparation, soutien des enfants <i>Q21</i>	Pension alimentaire ou revenu d'entretien (imposable) (<i>L128</i>)	Du côté fiscal, ces montants sont généralement payés en vertu d'une entente ou d'une décision d'un tribunal. Dans l'enquête, on s'appuie sur les mêmes définitions, et il est fait état de la L128.
Autres revenus monétaires imposables	Autres revenus monétaires imposables : sous-ensemble imposable de <i>Q17, Q23, Q24</i> et <i>Q25</i>	Autres revenus (<i>L130 + L104</i>)	Exemples : indemnité de cessation d'emploi, maintien du revenu, régime patronal ou syndical de prestations supplémentaires de chômage, paiements d'aide à l'enfance, bourses d'études, prestations de décès (enquête). Entrent également dans cette catégorie les autres revenus imposables de source gouvernementale non prévus explicitement ci-dessus.

6. Éléments non compris dans le revenu monétaire, mais requis dans le calcul du revenu imposable			
Retraits de REER	Retraits de REER <i>Q20</i>	Revenus d'un REER (rente ou retrait) si <65 ans <i>(L129)</i>	Retraits d'un REER non échu non comptés dans le revenu total, mais ajoutés au revenu imposable. Voir la note des revenus de pensions.
Gains en capital imposables	Gains en capital <i>Q6</i>	Gains en capital imposables <i>(L127)</i>	Les gains imposables équivalent à 75 % des gains nets. Les gains en capital imposables sont ajoutés au revenu imposable total, mais exclus du revenu monétaire total.
7. Autres éléments exclus			
Crédits d'impôt provinciaux	Crédits d'impôt provinciaux <i>Q15</i>	Crédits d'impôt provinciaux <i>(L479)</i>	Ces crédits sont exclus parce que les données de source fiscale ne sont pas disponibles dans certaines provinces. Du côté de l'enquête, de plus, il semble y avoir une sous-déclaration importante.
Pensions d'ancien combattant	Pensions d'ancien combattant <i>Q16</i>		Données recueillies dans l'enquête, mais non utilisées dans la fusion des données (ni dans les revenus totaux) parce qu'elles ne figurent pas dans les données fiscales.

2.3 Liens entre le travail et le revenu

De nombreux liens s'observent entre les données démographiques et les données sur le marché du travail recueillies dans le cadre de l'EDTR. L'interview avec rétro-information, dans laquelle des renseignements antérieurement recueillis auprès des répondants leur sont rappelés, aide à réduire les incompatibilités entre les données sur le travail déclarées en janvier et les données sur le revenu déclarées en mai. L'interview ne permet cependant pas d'éliminer toutes les incompatibilités, et il subsiste toujours des divergences. De plus, près des deux tiers des répondants ont déclaré leur revenu en donnant accès à leur dossier

fiscal. Il a été décidé au début de l'exploitation des données de ne pas corriger les données tirées des dossiers fiscaux. Dans certains cas, toutefois, les données sur le revenu recueillies dans le cadre des interviews ont été modifiées. L'inverse peut également se produire : les données sur le travail peuvent être corrigées si elles ne concordent pas avec des données sur le revenu qui paraissent correctes.

En janvier, on a demandé aux répondants de dire s'ils avaient reçu, à un moment de l'année, des indemnités pour accidents du travail, des prestations d'assurance-chômage ou des prestations d'assistance sociale. L'interview de janvier sert également à établir si une personne a travaillé comme employé ou comme travailleur indépendant pendant l'année. Les divers types de revenus devraient être rapportés dans l'interview de mai en fonction des réponses données. Par exemple, on devrait s'attendre à des rémunérations en mai si un répondant a déclaré en janvier avoir eu au moins un emploi de travailleur rémunéré. En fait, un indicateur a été ajouté en fonction de l'interview sur le travail et, si le revenu prévu n'était pas déclaré en mai, l'indicateur amenait une question exploratoire destinée à éliminer l'incompatibilité¹.

Comme on l'a vu ci-dessus, l'interview sur le travail comportait une question sur les mois pendant lesquels le répondant avait touché des prestations d'assistance sociale. Cette variable soulève une difficulté non seulement parce qu'il peut y avoir une incompatibilité entre les réponses données dans le cadre de l'interview sur le travail et la présence ou l'absence d'un revenu dans l'interview sur le revenu, mais en outre parce que les prestations d'assistance sociale sont une source de revenu familial. Elles peuvent être déclarées par une personne dans l'interview sur le travail et par une autre (habituellement un conjoint) dans l'interview sur le revenu (ou dans la déclaration d'impôt). Dans certaines familles, le

¹ Il s'agit là d'une «vérification interactive» que seul permet l'emploi de l'interview assistée par ordinateur.

montant déclaré est partagé. Bref, les possibilités de déclaration sont nombreuses, et la stratégie de vérification peut supposer des modifications mineures ou des modifications profondes. La première de ces deux stratégies a été retenue, car une intervention plus importante aurait supposé un trop grand nombre de décisions arbitraires. Les principes suivants ont été appliqués :

- Ne pas «obliger» un seul membre de la famille à déclarer la totalité des prestations d'assistance sociale de la famille. Ne pas répartir l'assistance sociale également entre tous les membres de la famille. Bref, laisser les données témoigner de la manière dont les familles choisissent de déclarer l'assistance sociale.
- Si un répondant a déclaré des prestations d'assistance sociale dans l'interview sur le travail, mais que d'autres membres de la famille ont déclaré des prestations dans l'interview sur le revenu, modifier l'interview sur le travail pour éliminer les divergences (ce qui ne change en rien les données sur la famille).
- Si un répondant a déclaré des prestations d'assistance sociale dans l'interview sur le revenu, mais qu'aucun membre de la famille n'en a déclaré dans l'interview sur le travail, modifier le statut du répondant dans l'interview sur le travail pour qu'il devienne «a reçu des prestations d'assistance sociale» (les mois pendant lesquels des prestations ont été touchées ne sont cependant pas imputés).
- Si un répondant a déclaré des prestations d'assistance sociale dans l'interview sur le travail, mais qu'aucun membre de la famille n'a déclaré de revenu à ce titre dans l'interview sur le revenu, imputer un montant au titre de l'assistance sociale. Cette procédure a été appliquée uniquement aux cas où les données sur les revenus ont été recueillies dans le cadre d'une interview.

L'interview sur le travail a également servi à déterminer la catégorie de travailleurs des répondants, selon qu'ils avaient travaillé pendant l'année comme employé, comme travailleur indépendant, etc. Le tableau ci-dessous présente certains des contrôles d'exploitation postérieurs à la collecte des données qui sont faits entre la catégorie de travailleurs déclarée dans l'interview sur le travail et le type de revenu prévu pour chaque répondant dans le cadre de l'interview sur le revenu.

TABLEAU 2. Compatibilité entre la catégorie de travailleurs et les renseignements sur le revenu

Si, dans l'interview sur le travail...	<i>...la valeur prévue se trouve sous la rubrique suivante :</i>
Des indemnités pour accidents du travail ont été reçues à un moment de l'année	<i>Indemnités pour accidents du travail (si le renseignement manque et que les données sur le revenu ont été obtenues par interview, un montant est imputé)</i>
Des prestations d'assurance-chômage ont été reçues à un moment de l'année	<i>Prestations d'assurance-chômage</i>
65 ans et plus, certains antécédents de travail et citoyen canadien depuis au moins dix ans	<i>Régime de pensions du Canada ou Régime de rentes du Québec</i>
Catégorie de travailleurs «employé» pour au moins un emploi	<i>Rémunérations</i>
Catégorie de travailleurs «travailleur familial non rémunéré» pour tous les emplois	<i>Gains nuls</i>
Catégorie de travailleurs «travailleur indépendant, entreprise constituée» pour au moins un emploi	<i>Rémunérations OU (et) revenus de placements</i>
Catégorie de travailleurs «travailleur indépendant, entreprise non constituée» pour au moins un emploi	<i>Revenu d'un travail indépendant agricole ou revenu d'un travail indépendant non agricole, compte tenu de l'activité économique</i>

L'approche retenue pour éliminer les incompatibilités que révèlent les comparaisons ci-dessus dépend de la provenance des données sur le revenu. Dans le cas des données recueillies par interview, si un montant manque ou paraît incompatible, il peut être imputé ou corrigé. Dans le cas des données sur le revenu tirées des dossiers fiscaux, on élimine si possible les incompatibilités en modifiant la catégorie de travailleurs, si la nature de l'erreur est évidente. Autrement, l'incompatibilité reste.

Les gains annuels totaux et le taux de rémunération horaire composite (données sur le travail) peuvent être imputés, s'ils manquent, à l'aide des rémunérations annuelles (données sur le revenu), pourvu que le répondant ait déclaré un seul emploi rémunéré et que le chiffre recueilli pour les rémunérations soit exact. Les gains annuels totaux équivalent au montant des rémunérations annuelles, et le taux de rémunération horaire composite est calculé en fonction des gains annuels totaux et des renseignements sur les heures travaillées. Il convient également de noter que les taux de rémunération et les gains annuels tirés des données de l'enquête sur le travail peuvent différer des rémunérations annuelles.

Quand d'autres renseignements sur le travail paraissent manquer — si, par exemple, un répondant a déclaré des rémunérations annuelles, mais pas d'emploi —, on ne fait pas d'imputation, car imputer des emplois et des caractéristiques d'emploi est une opération complexe. Ces incompatibilités subsistent donc dans les données, et il suffit de comparer les renseignements sur l'emploi aux données sur le revenu pour les relever.

2.4 Quelques principes fondamentaux de correction et d'imputation

Les données fiscales ne sont pas modifiées

Les données de près de la moitié des répondants de la première phase viennent des fichiers d'impôt. Il a été décidé de ne pas modifier les données fiscales si des incompatibilités avec des données sur le travail ou des données démographiques étaient observées. Trois raisons ont motivé cette décision :

- Il est difficile de corriger «superficiellement» des données sur le revenu en raison de la relation qu'il y a entre les divers éléments de revenu. Par exemple, si une source de revenu est modifiée, le revenu total et l'impôt exigible s'en ressentent. En ce sens, toute intervention a un effet d'entraînement.
- Les corrections des données fiscales sur le revenu auraient suscité des incompatibilités longitudinales.
- Modifier les données fiscales aurait compliqué un système d'exploitation déjà complexe.

En rétrospective, il paraît raisonnable de faire quelques contrôles de base par mesure de précaution, et quelques-uns seront probablement adoptés pour l'exploitation des données de la deuxième phase.

Les écarts entre les rémunérations et les gains totaux tirés d'un emploi rémunéré sont tolérés

Dans le cadre de l'interview sur le travail, on recueille des renseignements sur tous les emplois rémunérés des répondants afin de calculer le taux de rémunération horaire et les gains annuels totaux de chaque emploi. La somme des gains annuels totaux de la totalité des emplois rémunérés équivaut à peu près, théoriquement, aux renseignements recueillis sur les rémunérations dans le cadre de l'interview sur le revenu. Nous avons décidé de ne pas chercher à éliminer les écarts entre les deux par correction et imputation. C'est la première fois que nous sommes en mesure de faire une micro-analyse des variations des régimes de déclaration qui s'observent dans ces deux modèles bien établis de collecte de données sur les gains. Quand les régimes auront été bien compris, on pourra peut-être envisager de faire des corrections.

3. CORRECTION DES DONNÉES D'ENQUÊTE

Les données recueillies ont été soumises à un certain nombre de corrections. Cette section présente une description des mesures prises pour épurer les données d'enquête avant de les fusionner aux données fiscales.

3.1 Classement des observations des intervieweurs et réaction

Le traitement des observations des intervieweurs s'est fait en deux étapes : codage des observations et réaction aux observations. Les observations qui concernaient directement le revenu ont été codées comme telles à l'aide d'un programme de codage spécial.

En réaction aux observations sur le revenu, on a établi une liste de mesures à prendre à l'égard de divers types d'observations et on l'a utilisée pour corriger le fichier des revenus.

3.2 Comparaison du revenu total indiqué (Q26) à la somme des éléments (Q01 à Q25)

L'étape suivante a consisté à comparer le revenu total indiqué (Q26) à la somme des éléments (Q01 à Q25) dans le but de vérifier la qualité des données. Il faut en effet garder à l'esprit que le revenu total **indiqué** (Q26) ne correspond pas nécessairement au revenu total **déclaré** par le répondant, mais qu'il peut représenter plutôt le revenu **calculé** par l'ordinateur.

Quand un intervieweur entre des montants pendant l'interview sur le revenu, l'ordinateur fait un total cumulatif de ces montants. Une fois que tous les revenus du répondant ont été indiqués, l'intervieweur demande au répondant si le total cumulatif semble correct. Si le répondant estime que le total n'est pas exact, l'intervieweur peut indiquer un chiffre différent. En général, les seuls cas où le revenu total et la somme des éléments ne sont **pas** égaux sont ceux où le répondant a donné des réponses «ne sait pas» ou a refusé de répondre à certaines questions ou qu'il n'a pas donné de montant, si ce n'est pour le revenu total.

Il convient également de noter que, chaque fois que le revenu total a été calculé par l'ordinateur, le montant a été arrondi au dollar. Ainsi, pour faire nos comparaisons entre le revenu total indiqué et la somme des éléments, nous avons postulé que les deux étaient égaux si l'écart était inférieur à 2 \$. Les résultats obtenus sont présentés au tableau 3; ils révèlent que, dans 91,3 % des cas (115+233+12+20 116), les données semblent correctes. Ainsi, le revenu total indiqué — ou calculé par l'ordinateur — est exactement égal à la somme des éléments dans 89,7 % des cas (20 116 cas).

Tableau 3. Comparaison du revenu total indiqué en Q26 et de la somme des éléments (Q01 à Q25), dans les cas où le revenu total est >0

Types de réponse aux questions Q01 à Q25 prises ensemble	Résultats de la comparaison			
	Somme <Q26	Somme = Q26	Somme >Q26	Total
Aucun revenu d'aucune forme pour toutes les questions	115 0.5%	12 0.1%	0 0.0%	127 0.6%
Oui à certaines questions**/non à certaines questions	352 1.6%	20 116 89.7%	491 2.2%	20 959 93.4%
Oui / non / ne sait pas ou refus	99 0.4%	997 4.4%	18 0.1%	1 114 5.0%
Non / ne sait pas ou refus	233 1.0%	0 0.0%	0 0.0%	233 1.0%
Total	799 3.5%	21 125 94.2%	509 2.3%	22 433 100.0%

* 0 \$ ou blanc ** montant

Pour essayer de déterminer comment les diverses formes d'erreur devraient être corrigées, on a procédé à un contrôle manuel. Dans certains cas, le revenu total ou un élément du revenu a été modifié, compte tenu de la logique de la situation. Dans d'autres cas, le revenu total a été laissé inchangé, puisqu'il était impossible de savoir lequel des deux montants était correct (la somme des éléments ou le revenu total). Dans tous les cas, le revenu total a été recalculé par totalisation des éléments chaque fois qu'une imputation a été faite.

3.3 Vérifications

3.3.1 Prestation fiscale pour enfants

Pour recevoir une prestation fiscale pour enfants, une personne doit être responsable d'un enfant de moins de 18 ans. Comme nous avons craint que des personnes âgées aient déclaré par erreur qu'elles touchaient une PFE, nous avons examiné le dossier de toutes les personnes de 60 ans et plus qui avaient déclaré un montant en Q08 (prestation fiscale pour enfants). Trois cas ont été relevés, et chacune de ces personnes avaient effectivement un enfant.

3.3.2 Pension de sécurité de la vieillesse, supplément de revenu garanti et allocation de conjoint

En général, toute personne de 65 ans et plus devrait recevoir la pension de sécurité de la vieillesse (sauf si elle a immigré il y a moins de 10 ans ou que son revenu à la retraite dépasse 80 000 \$). Comme les conditions d'admissibilité à la PSV sont connues, il a été possible de repérer les personnes qui n'avaient pas déclaré de PSV, mais qui, selon toute vraisemblance, l'avaient touchée. Nous avons constaté que certaines personnes de 65 ans et plus n'avaient pas déclaré de pension de sécurité de la vieillesse ou de supplément de revenu garanti (Q09). Pourquoi en était-il ainsi?

Nous avons tout d'abord comparé les données d'enquête et les données fiscales des personnes de 65 ans et plus pour lesquelles Q09 = 0. En général, nous avons constaté que le montant qui aurait dû être déclaré en Q09 l'avait été ailleurs dans l'enquête ou avait été ajouté au montant donné en Q10 (Régime de pensions du Canada ou Régime de rentes du Québec). Si le montant exact qui aurait dû figurer en Q09 (4 586,16 \$ en 1993, par

exemple) ou un montant voisin figurait ailleurs qu'en Q09, il a été déplacé. Nous avons en outre décidé de donner, dans le manuel de l'intervieweur, des renseignements plus précis sur les pensions des personnes âgées. Nous espérons que cela contribuera à réduire le nombre des erreurs.

3.3.3 Pension de l'employeur

Le tableau 4 présente une comparaison des répondants de 59 ans et plus qui n'ont pas indiqué de montant en Q18 et de ce qui figurait dans leur déclaration d'impôt.

Tableau 4. Comparaison des données fiscales et des données d'enquête au titre des pensions d'employeur (Q18) des personnes de 59 ans et plus

Enquête (Q18)	déclaration d'impôt		
	montant	zéro	total
ne sait pas	54 2.9%	5 0.3%	59 3.1%
refus	27 1.4%	8 0.4%	35 1.9%
zéro	437 23.3%	1 345 71.7%	1 782 95.0%
total	518 27.6%	1 358 72.4%	1 876 100.0%

Des 1 876 répondants qui se trouvaient dans cette situation, 28 % ont déclaré ne pas toucher de pension d'employeur dans l'enquête, mais ont indiqué un montant dans leur

déclaration d'impôt. Le même tableau établi pour des groupes d'âge plus fins révèle que :

- 16,8 % des personnes de 59 à 64 ans auraient dû déclarer un montant en Q18, mais ne l'ont pas fait;
- 26,7 % des personnes de 65 à 69 ans auraient dû déclarer un montant en Q18, mais ne l'ont pas fait;
- 24,8 % des personnes de 70 ans et plus auraient dû déclarer un montant en Q18, mais ne l'ont pas fait.

3.3.4 Régime de pensions du Canada/Régime de rentes du Québec

Pour évaluer la qualité de la déclaration des revenus de pension, nous avons comparé les montants donnés en réponse à cette question dans l'interview sur le revenu aux données des fichiers d'impôt. Nous avons constaté que la réponse à la Q10 manquait souvent dans l'enquête. Les tableaux 5 à 8 donnent un aperçu du problème.

Tableau 5. Comparaison des données fiscales/données d'enquête au titre du Régime de pensions du Canada et du Régime de rentes du Québec (Q10), tous âges

Enquête	déclaration d'impôt		
	montant	zéro	total
ne sait pas/refus	244 1.2%	68 0.3%	312 1.5%
montant	3 057 13.9%	107 0.6%	3 164 14.5%
zéro	682 3.3%	17 080 80.7%	17 762 84.0%
total	3 983 18.4%	17 255 81.6%	21 238 100.0%

Le tableau 5 révèle que 4,5 % des membres de la population observée (244+682) n'ont pas indiqué de montant dans l'enquête, alors qu'ils avaient déclaré recevoir une rente du RPC/RRQ dans leur déclaration d'impôt. Nous avons étudié de plus près les données fiscales des personnes de 60 ans et plus du point de vue des antécédents de travail.

Tableau 6. Présence ou absence d'un montant au titre du RPC/RRQ dans la déclaration d'impôt, selon les antécédents de travail, personnes de 60 à 64 ans

RPC/RRQ (données fiscales)	antécédents de travail (données d'enquête)		
	ne sait pas	avec antécédents	sans antécédents
montant	97 32,9 %	405 79,3 %	11 44 %
zéro	198 67,1 %	106 20,7 %	14 56 %
total	295 100 %	511 100 %	25 100 %

Tableau 7. Présence ou absence d'un montant au titre du RPC/RRQ dans la déclaration d'impôt, selon les antécédents de travail, personnes de 65 à 69 ans

RPC/RRQ (données fiscales)	antécédents de travail		
	ne sait pas	avec antécédents	sans antécédents
montant	140 91,5 %	786 91,5 %	28 47,5 %
zéro	13 8,5 %	73 8,5 %	31 52,5 %
total	153 100 %	859 100 %	59 100 %

Tableau 8. Présence ou absence d'un montant au titre du RPC/RRQ dans la déclaration d'impôt, selon les antécédents de travail, personnes de 70 ans et plus

RPC/RRQ (données fiscales)	antécédents de travail		
	ne sait pas	avec antécédents	sans antécédents
montant	92 93,9 %	1 577 86 %	156 59,3 %
zéro	6 6,1 %	256 14 %	107 40,7 %
total	98 100 %	1 833 100 %	263 100 %

Les trois tableaux ci-dessus montrent que, selon les données fiscales, 79 % des répondants de 60 à 64 ans qui avaient des antécédents de travail recevaient une rente du RPC/RRQ. Chez les personnes de 65 à 69 ans, cette proportion passait à 92 %. Chez les personnes de 70 ans et plus, elle s'établissait à 86 %, sans doute parce que certaines d'entre elles n'avaient pas travaillé depuis l'adoption de la législation sur le RPC, en 1966.

Règles d'attribution de l'indicateur RPC/RRQ

Compte tenu de la sous-déclaration des rentes du RPC/RRQ, on a décidé de faire une imputation quand une rente avait une forte probabilité d'être reçue. La méthode adoptée et son utilité sont succinctement exposées ci-dessous. Les règles établies pour les groupes d'âge de 65 à 69 ans et de 70 ans et plus sont les mêmes et elles sont simples; les règles qui s'appliquent aux personnes de 60 à 64 ans sont plus complexes.

Tableau 9. Comparaison des données d'enquête et des données fiscales au titre de la déclaration de rentes du RPC/RRQ, personnes de 70 ans et plus

DONNÉES D'ENQUÊTE	DONNÉES FISCALES		
	RPC=0	RPC=montant	Total
RPC=0	303 14,8 %	219 10,7 %	522 25,4 %
RPC=ne sait pas ou refus	15 0,7 %	134 6,5 %	149 7,3 %
RPC=montant	33 1,6 %	1 349 65,7 %	1 382 67,3 %
Total	351 17,1 %	1 702 82,9 %	2 053 100 %

Le tableau 9 montre que, s'il n'avait pas été fait de vérification, 353 répondants de 70 ans et plus auraient été rangés par erreur parmi les personnes ne touchant pas de rente du RPC/RRQ. Ce chiffre représente 17,2 % de la population observée ou 52,6 % des répondants de ce groupe d'âge ayant déclaré à l'enquête ne pas toucher de rente du RPC/RRQ.

Pour corriger cette situation, nous avons assimilé le cas de personnes de ce groupe d'âge qui avaient travaillé depuis 1966 à une non réponse chaque fois qu'une rente du RPC/RRQ n'était pas déclarée, et nous avons imputé un montant à partir d'un dossier analogue.

Cette vérification était justifiée. Le tableau 10 montre que l'approche retenue a mené à une mesure qui était justifiée dans 492 cas, mais non dans 179 cas (ce total de 671 cas représente la totalité des personnes de 70 ans et plus qui, à tort ou à raison, n'avaient pas déclaré de rente du RPC/RRQ à l'interview).

Tableau 10. Erreurs causées par l'algorithme d'imputation, personnes de 70 ans et plus

IMPUTATION	ERREUR		
	non	oui	Total
non	214 31,9 %	75 11,2 %	289 43,1 %
oui	278 41,4 %	104 15,5 %	382 56,9 %
Total	492 73,3 %	179 26,7 %	671 100 %

Tableau 11. Comparaison des données d'enquête et des données fiscales au titre de la déclaration de rentes du RPC/RRQ, personnes de 65 à 69 ans

DONNÉES D'ENQUÊTE	DONNÉES FISCALES		
	RPC=0	RPC=montant	Total
RPC=0	102 10,5 %	99 11,5 %	201 20,6 %
RPC=ne sait pas ou refus	3 2,7 %	63 7,3 %	66 6,8 %
RPC=montant	6 0,6 %	702 72 %	1 708 72,6 %
Total	111 11,4 %	864 88,6 %	975 100 %

À la lecture du tableau 11, on constate que 162 répondants de 65 à 69 ans auraient pu être rangés par erreur parmi les personnes ne touchant pas de rente du RPC/RRQ, ce chiffre correspondant à 60,6 % des répondants n'ayant pas déclaré de revenu de cette source.

La méthode de correction retenue dans ce cas était semblable à celle qui a été appliquée aux personnes de 70 ans et plus. L'effet en est présenté au tableau 12. Essentiellement, la correction ne s'est révélée injustifiée que dans 18,7 % des cas.

Tableau 12. Erreurs causées par l'algorithme d'imputation, personnes de 65 à 69 ans

IMPUTATION	ERREUR		
	non	oui	Total
non	69 25,8 %	14 5,2 %	83 31,1 %
oui	148 55,4 %	36 13,5 %	184 68,9 %
Total	217 81,3 %	50 18,7 %	267 100 %

Tableau 13. Comparaison des données d'enquête et des données fiscales au titre de la déclaration de rentes du RPC/RRQ, personnes de 60 à 64 ans

DONNÉES D'ENQUÊTE	DONNÉES FISCALES		
	RPC=0	RPC=montant	Total
RPC=0	270 34,4 %	99 12,6 %	369 47,0 %
RPC=ne sait pas ou refus	4 0,5 %	21 2,7 %	25 3,2 %
RPC=montant	7 0,9 %	385 49 %	392 49,9 %
Total	281 35,8 %	505 64,3 %	786 100 %

Le nombre de répondants de 60 à 64 ans qui auraient par erreur été rangés parmi les personnes ne touchant pas de rente du RPC/RRQ aurait été de 120 en l'absence de vérification, ce qui représente 30,5 % des répondants n'ayant déclaré aucun montant dans l'enquête. Comme nous l'avons souligné ci-dessus, l'algorithme utilisé pour établir l'opportunité de faire une imputation de rente du RPC/RRQ pour une personne de ce groupe d'âge était un peu plus complexe. Ainsi, il n'était pas fait d'imputation si, à en juger par les antécédents de travail, une personne travaillait toujours. De même, il n'a pas été fait d'imputation pour les personnes qui avaient arrêté de travailler avant 1967 (c'est-à-dire à l'époque de l'adoption du régime) ou qui avaient arrêté de travailler après 1966, mais qui avaient perdu leur emploi ou qui avaient été mises à pied. Par contre, une imputation a été faite pour les personnes qui ont cessé de travailler après 1966 pour d'autres raisons.

Tableau 14. Erreurs causées par l'algorithme d'imputation, personnes de 60 à 64 ans

IMPUTATION	ERREUR		
	non	oui	Total
non	231 56,6 %	41 10,4 %	272 69 %
oui	79 20,1 %	43 10,9 %	122 31 %
Total	310 78,7 %	84 21,3 %	394 100 %

Le tableau 14 révèle que cette méthode a entraîné des erreurs de classement dans 21 % des cas de non-réponse à l'enquête.

3.3.5 Cas «à préciser»

La première étape du traitement des cas «à préciser» consiste à les coder. Les dossiers sont tout d'abord codés automatiquement à l'aide d'un programme de codage automatisé de l'EDTR, en fonction d'un fichier dictionnaire de l'Enquête sur les finances des consommateurs. Si un code n'est pas attribué automatiquement, le codage se fait manuellement. Une fois que tous les cas «à préciser» ont été codés, les montants sont reportés dans la catégorie de revenu appropriée, à l'aide d'un programme qui s'appuie sur des spécifications de l'EFC. Pour appliquer cette mesure, nous avons dû recourir à la variable «catégorie de travailleurs». Des spécifications d'obtention de la variable «catégorie de travailleurs» ont été définies de manière que cette opération puisse être menée à bien avant la fin de l'exploitation des données sur le travail. (Il s'agissait là d'une mesure à court terme permettant une exploitation parallèle des données sur le travail et sur le revenu de la première phase; dans la deuxième phase, cette étape de l'exploitation des données sur le revenu se fera après le codage de la catégorie de travailleurs.)

3.4 Étude des valeurs extrêmes

Une définition des «valeurs extrêmes» a été énoncée en fonction des données fiscales des répondants de l'enquête. Les intervalles obtenus (après arrondissement) pour chaque élément de revenu sont présentés au tableau 15.

**Tableau 15. Intervalles de revenu de la population observée, par source de revenu
(données fiscales)**

Source de revenu	Intervalle observé	
	Minimum	Maximum
Rémunérations	0 \$	1 500 000 \$
Revenu net d'un travail agricole indépendant	-65 000 \$	90 000 \$
Revenu net d'un travail indépendant non agricole	-1 050 000 \$	740 000 \$
Revenu de placements (int., div. et autres rev. de placements)	-190 000 \$	300 000 \$
Gains en capital nets	0 \$	630 000 \$
Prestation fiscale pour enfants	0 \$	9 600 \$
PSV/SRG/AC	0 \$	19 000 \$
RPC/RRQ	0 \$	23 000 \$
Assurance-chômage	0 \$	23 000 \$
Assistance sociale	0 \$	25 000 \$
Indemnités pour accidents du travail	0 \$	59 000 \$
Crédit d'impôt provincial	0 \$	7 500 \$
Pensions de l'employeur, rentes de REER, retraits de FERR	0 \$	110 000 \$
Retraits de REER	0 \$	78 000 \$
Autres revenus monétaires (imposables)	0 \$	200 000 \$

Les données sur le revenu recueillies par interview qui ne tombaient pas dans ces intervalles ont été examinées. Seize cas ont été observés, dans 7 catégories de revenu.

Ces cas particuliers peuvent vraisemblablement être assimilés à des erreurs; ils sont décrits au tableau 16.

Tableau 16. Observations faites dans les interviews sur le revenu qui tombent à l'extérieur des intervalles déterminés avec les données fiscales

Source de revenu	Nature/montant
Rémunérations	- 1 montant négatif (moins de -2 000 \$) - 1 montant positif extrême (plus de 3 millions de dollars)
Revenu d'un travail agricole indépendant	- 4 montants de plus de 100 000 \$
Gains en capital nets	- 4 montants négatifs de -900 \$ à -15 000 \$
PFE	- 2 observations dépassant le maximum des fichiers d'impôt (plus de 10 000 \$)
RPC/RRQ	- 2 observations dépassant le maximum
Assurance-chômage	- 1 observation dépassant le maximum
Crédit d'impôt provincial	- 1 observation dépassant le maximum

Dans le fichier à grande diffusion de la première phase, les trois revenus les plus élevés de chacune des sources ont été supprimés, ce qui a permis d'éliminer la plupart des cas particuliers. Dans la deuxième phase, la méthode de détection des cas particuliers permettra de faire une distinction entre les valeurs extrêmes qui sont manifestement erronées et celles qui sont valides. Seules les valeurs de la première catégorie feront l'objet d'une vérification.

3.5 Attribution de codes-réponses aux données sur le revenu

La méthode d'attribution de codes-réponses aux répondants dépend en partie de la provenance des données : fichier d'impôt ou interview. Si les données sur le revenu sont tirées du fichier d'impôt, le code-réponse du répondant est «complet» (code 000), à moins que l'un des champs des données fiscales du répondant contienne un «1», c'est-à-dire un

code d'exploitation et non un montant en dollars. Dans ce dernier cas, le code-réponse devient «partiel» (code 001). Un «1» a été relevé dans au moins un champ dans 139 cas. Ces cas ont été examinés un à un. Dans 85 cas, le code concernait les rémunérations. Dans 28 cas, il concernait les revenus de placements, et dans 14, la PSV. Les autres cas se répartissaient parmi les diverses catégories de revenu. Les observations de type «1» ont été éliminées pendant l'exploitation des données.

Quand les renseignements avaient été obtenus par interview, par contre, le code-réponse du répondant était fonction de la nature des réponses obtenues à l'interview. Deux variables ont été créées à cette fin : **RESP25** décrit la réponse obtenue aux questions sur les éléments du revenu (Q1 à Q25) et **Q26** indique la réponse obtenue à la question sur le revenu total (Q26).

RESP25 peut prendre l'une des cinq valeurs suivantes :

1	Non	0 \$ ou blanc pour toutes les sources de revenu
2	Oui/non	montant inscrit dans au moins un élément + 0 \$ ou blanc pour certains éléments
3	Oui/non/ne sait pas ou refus	au moins un montant + 0 \$ ou blanc pour certains éléments + au moins un élément marqué «ne sait pas» ou «refus»
4	Non/ne sait pas ou refus	0 \$ ou blanc dans certains éléments et reste des éléments marqués «ne sait pas» ou «refus»
5	Ne sait pas ou refus	Tous les éléments marqués «ne sait pas» ou «refus»

Q26 peut prendre l'une des trois valeurs suivantes :

1	Non	0 \$ ou blanc pour le revenu total
2	Oui	montant indiqué pour le revenu total
3	Ne sait pas ou refus	ne sait pas ou refus indiqué pour le revenu total

Un code-réponse a été défini pour chaque répondant en fonction de ces deux variables.

Une **réponse complète** correspond à l'une des combinaisons suivantes :

	Q26	RESP25 et code-réponse de collecte
1	Oui	Oui/non
2	Non	Non et code-réponse=000 (complet)*
3	Non	Oui/non
4	Ne sait pas/refus	Oui/non

* Correspond à une personne sans revenu.

La **non-réponse totale** sert à désigner les personnes qui n'ont déclaré aucun revenu, mais qui ne sont pas «sans revenu». La non-réponse totale peut prendre l'une des formes suivantes :

	Q26	RESP25 et code-réponse de collecte
1	Non	Non et code-réponse=non-réponse
2	Refus	Non
	Q26	RESP25 et code-réponse de collecte
3	Ne sait pas	Non
4	Non	Ne sait pas/refus
5	Refus	Ne sait pas/refus
6	Ne sait pas	Ne sait pas/refus
7	Non	Non/ne sait pas/refus
8	Refus	Non/ne sait pas/refus
9	Ne sait pas	Non/ne sait pas/refus
10	Non	Oui/non et code-réponse=non-réponse
11	Refus	Oui/non et code-réponse=non-réponse
12	Ne sait pas	Oui/non et code-réponse=non-réponse
13	Non	Oui/non/refus/ne sait pas et code-réponse=non-réponse
14	Refus	Oui/non/refus/ne sait pas et code-réponse=non-réponse
15	Ne sait pas	Oui/non/refus/ne sait pas et code-réponse=non-réponse

Les autres combinaisons ont toutes été assimilées à une **réponse partielle**.

3.6 Compatibilité entre la catégorie de travailleurs et le revenu d'emploi

Dans l'interview sur le travail, un répondant peut déclarer avoir eu jusqu'à six emplois. Compte tenu de la catégorie de travailleurs, des rémunérations ou un revenu d'un travail indépendant devraient être déclarés dans l'interview sur le revenu. Pour être plus précis :

- si la catégorie de travailleurs est *travailleur rémunéré* (COW=1) dans au moins un des emplois, le revenu d'emploi devrait être déclaré en Q01 (rémunérations);

- si la catégorie de travailleurs est *travailleur indépendant, constitué en corporation* (COW=3 ou 4) dans au moins un des emplois, le revenu d'emploi devrait être déclaré en Q01 (rémunérations) ou en Q05 (dividendes);
- si la catégorie de travailleurs est *travailleur familial non rémunéré* (COW=2), aucun revenu d'emploi ne devrait être déclaré;
- si la catégorie de travailleurs est *travailleur indépendant, non constitué en corporation* (COW=5 ou 6), le revenu d'emploi devrait être déclaré en Q02 (revenu net d'un travail agricole indépendant) ou en Q03 (revenu net d'un travail indépendant non agricole), compte tenu du code d'activité économique.

Pour ce faire, nous sommes partis de la variable «catégorie de travailleurs» utilisée dans le report des cas «à préciser» à l'étape 3.3.5. Nous avons établi des règles pour comparer la catégorie de travailleurs en fonction des renseignements recueillis dans l'interview sur le travail et du revenu déclaré dans l'interview sur le revenu (tableau 17). En cas d'incompatibilité entre le travail et le revenu, deux solutions s'offrent selon que les données viennent de l'enquête ou des fichiers d'impôt : on peut soit corriger l'incompatibilité en faisant passer le revenu à une autre catégorie ou en modifiant la catégorie de travailleurs dans l'interview sur le travail, soit ne rien modifier et attribuer aux dossiers un indicateur d'incompatibilité.

Tableau 17. Chiffres sur les mesures prises pour éliminer des incompatibilités entre les renseignements sur le travail et sur le revenu

	Chiffres		
	Total	Enquête	Impôt
Nombre total de dossiers	23 030	9 517	13 513
Pas d'incompatibilité entre le travail et le revenu	20 229	8 334	11 895
Certaines incompatibilités, mais aucune mesure prise*	1 785	602	1 183
Incompatibilités à éliminer par modification d'au moins une catégorie de travailleurs	602	167	435
Montant déplacé d'un revenu d'un travail indépendant agricole à un travail indépendant non agricole ou vice versa	116	116	0
Montant déplacé d'un revenu d'un travail indépendant non agricole à des rémunérations	57	57	0
Montant imputé par le SGVI : rémunérations, revenu d'un travail indépendant ou revenus de placements	241	241	0

* Ces dossiers seront exclus du bassin des donneurs dans les imputations (SGVI).

** Écart entre les données vérifiées et les données recueillies au sujet de la catégorie de travailleurs.

4. APPARIEMENT DES FICHIERS D'IMPÔT

Quand les données sur le revenu de la première phase ont été recueillies, en mai 1994, l'opportunité d'offrir aux répondants la possibilité d'autoriser l'accès à leur dossier fiscal n'avait pas encore été évaluée. Cette évaluation était néanmoins terminée au début de 1995. On a alors décidé d'offrir cette possibilité aux répondants à l'occasion l'interview de

mai 1995 (les années suivantes, l'offre devait se faire en janvier, de manière qu'il ne soit pas nécessaire de communiquer en mai avec les ménages dont tous les membres auraient donné accès à leur dossier fiscal).

En mai 1995, on a donc demandé aux répondants la permission d'utiliser des données fiscales pour toute la durée de leur participation à l'enquête. Il a été décidé qu'on utiliserait les données fiscales de 1993 et des années subséquentes des répondants qui avaient donné leur accord et dont le dossier fiscal a pu être apparié. Cette mesure visait principalement à faire en sorte que le changement de source de données n'ait pas d'effet sur les micro-variations de 1993 à 1994.

En mai 1995, environ 62 % des répondants ont donné accès à leur dossier fiscal. Les données fiscales ont été utilisées dans tous les cas, sauf quand il a été impossible de trouver un dossier fiscal correspondant. Dans ce cas, les résultats de l'interview sur le revenu ont été utilisés. Il convient de noter que, quand la possibilité de donner accès au dossier fiscal a de nouveau été offerte en janvier 1996 (aux nouveaux répondants et à ceux qui avaient refusé la fois précédente), la proportion de ceux qui ont accepté est passée à près de 73 %.

4.1 Description sommaire de la méthode utilisée : appariement direct ou appariement statistique

Dans l'EDTR, on ne demande pas aux répondants leur NAS, car il s'agit là d'un renseignement délicat à solliciter par téléphone et que cela soulève des difficultés. L'appariement se fait plutôt par le biais d'une procédure statistique en deux étapes.

La première étape consiste en un appariement direct du fichier d'impôt en fonction du

nom, de la date de naissance, du sexe, de l'état matrimonial et du code postal. Un appariement a pu être fait dans 50 % des cas à l'aide de cette méthode. La deuxième étape consiste en un appariement statistique dans lequel on admet que certaines des valeurs d'appariement manquent. Cette mesure a permis de faire passer à 85 % la proportion des dossiers appariés.

4.2 Quelques taux d'appariement et plans d'évaluation

Le tableau 18 présente le nombre total de répondants de l'EDTR dont le dossier fiscal n'a pas été utilisé pour 1993, pour l'une des raisons suivantes :

- le répondant est un non-déclarant;
- le répondant n'a pas autorisé l'accès à son dossier fiscal;
- le répondant a donné accès à son dossier fiscal, mais le Bureau n'a pas réussi à faire d'appariement.

Un ménage peut avoir trois statuts de code-réponse : complet (ce qui signifie que tous les membres admissibles du ménage ont participé à une interview complète), partiel et non-réponse. Chaque ligne correspond à une paire de codes-réponses : le premier est le code-réponse de l'interview de janvier sur le travail, et le second, le code-réponse de l'interview de mai sur le revenu. Par exemple, «complet/complet» signifie que des interviews complètes ont été faites auprès de tous les membres d'un ménage en janvier et en mai.

Dans le tableau 18, les cas nécessitant une imputation ont été signalés. L'imputation a été faite avec le Système généralisé de vérification et d'imputation (SGVI). Chaque fois que cela était possible, un répondant ayant donné accès à son dossier fiscal était choisi comme

donneur. Le tableau 19 présente des résultats comparables sur les répondants dont les données sur le revenu de 1993 ont été tirées des fichiers d'impôt.

Points saillants

- Le nombre de personnes dont le dossier fiscal a été utilisé dans la première phase était légèrement inférieur à 15 000, ce qui représentait environ la moitié des répondants de 16 ans et plus.
- Dans le tableau 18, le modèle le plus répandu, outre le modèle «complet/complet» est le modèle «complet/non-réponse». Il s'agit de cas (3 311) dans lesquels on a pu faire une interview complète au sujet du travail, mais pour lesquels il n'a pas été possible d'obtenir de renseignements sur le revenu.

Tableau 18. Code-réponse : personnes pour lesquelles des données d'enquête ont été utilisées dans la première phase

Code-réponse janv./mai	Total	Répondants longitudinaux	Cohabitants
Total	15 055	14 180	875
Complet/complet	9 446	9 072	374
Complet/partiel	842	787*	55*
Partiel/complet	103	101	2
Partiel/partiel	1	1*	0
Complet/non-réponse	3 311	3 092**	219**
Partiel/non-réponse	31	31**	0
Non-réponse/complet	770	751	19
Non-réponse/partiel	62	54*	8*
Non-réponse/non-réponse	489	291**	198**

* Imputation partielle

** Imputation complète

Tableau 19. Code-réponse : personnes dont on a utilisé le dossier fiscal (accès autorisé et appariement réussi)

Code-réponse janv./mai	Total	Répondants longitudinaux	Cohabitants
Total	14 627	14 082	545
Complet/complet	13 318	12 868	450
Complet/partiel	125	122	3
Partiel/complet	14	13	1
Partiel/partiel	1	1	0
Complet/non-réponse	0	0	0
Partiel/non-réponse	0	0	0
Non-réponse/complet	1 161	1 071	90
Non-réponse/partiel	8	7	1
Non-réponse/non-réponse	0	0	0

Les tableaux qui suivent montrent où il a fallu imputer des revenus de diverses sources, selon que les répondants ont donné ou non accès à leur fichier d'impôt. Le premier de chacune des paires de tableau concerne des répondants pour lesquels on a utilisé les données d'enquête et non les données fiscales. Outre les personnes qui n'ont pas donné accès à leur dossier fiscal, ces répondants comprennent les non-déclarants et les cas de non-appariement.

Quelques explications au sujet des termes utilisés

Indicateur en janvier signifie que, compte tenu des résultats de l'interview sur le travail de janvier, un revenu devrait être déclaré pour la source en question. En d'autres termes, nous nous attendions à ce que le répondant déclare des rémunérations dans l'interview de mai puisqu'il avait déclaré au moins un emploi rémunéré en janvier. Chaque fois qu'un

«indicateur» était ainsi mis et qu'aucun revenu n'était déclaré en mai, l'intervieweur était amené par l'indicateur à poser une question exploratoire en vue de dissiper l'incompatibilité. Les cas présentés ci-dessous sont ceux où subsiste une incompatibilité. Dans les tableaux sur les répondants qui ont donné accès à leur dossier fiscal (et dont le fichier d'impôt a pu être apparié), l'«indicateur» indique simplement qu'une source de revenu en particulier devrait être déclarée, compte tenu de l'interview sur le travail.

Quand une incompatibilité est relevée entre les données sur le travail et les données sur le revenu, des mesures peuvent être prises ou non. Quand une mesure est prise, celle-ci consiste soit à imputer un montant (* dans les tableaux) ou à modifier l'indicateur (** dans les tableaux).

Tableau 20. Répondants ayant déclaré des rémunérations — Données d'enquête

Code-réponse janv./mai	Total	Pas d'indicateur en janv. et pas de montant en mai	Indicateur en janv., mais pas de montant en mai	Non- réponse en janv.	Pas d'indicateur en janv., mais montant en mai	Indicateur en janv. et montant en mai	Autre
Total	15 055	5 869	2 207	765	556	5 196	462
Comp/comp	9 446	3 903	169*	2	475**	4 897	0
Comp/part	842	359	160	0	29**	294	0
Part/comp	103	63	0	0	37**	3	0
Part/part	1	1	0	0	0	0	0
Comp/non	3 311	1 432	1 876	1	0	2	0
Part/non	31	28	2	1	0	0	0
Non/comp	770	7	0	302	15**	0	446
Non/part	62	0	0	47	0	0	15
Non/non	489	76	0	412	0	0	1

Tableau 21. Répondants ayant déclaré des rémunérations — Données fiscales

Code-réponse janv./mai	Total	Pas d'indicateur en janv. et pas de montant en mai	Indicateur en janv., mais pas de montant en mai	Non- réponse en janv.	Pas d'indicateur en janv., mais montant en mai	Indicateur en janv. et montant en mai	Autre
Total	14 627	4 122	263	355	728	8 370	789
Comp/comp	13 318	4 028	253	1	692**	8 340	4
Comp/part	125	86	10	0	2**	27	0
Part/comp	14	3	0	0	8**	3	0
Part/part	1	0	0	0	1**	0	0
Comp/non	0	0	0	0	0	0	0
Part/non	0	0	0	0	0	0	0
Non/comp	1 161	5	0	347	25	0	784
Non/part	8	0	0	7	0	0	1
Non/non	0	0	0	0	0	0	0

* Imputer le montant

** Changer l'indicateur

La première rangée du tableau 21 présente les mêmes renseignements pour les répondants qui ont donné accès à leur dossier fiscal. Moins de 2 % des répondants qui ont déclaré avoir un emploi en janvier n'avaient pas de rémunérations dans leurs revenus. En revanche, 728 (5 %) répondants n'ont pas fait état d'un emploi rémunéré dans l'interview sur le travail, mais ont déclaré des rémunérations dans leur déclaration d'impôt.

Ainsi, on observe moins souvent d'écart «travail/revenu» au titre de la déclaration des rémunérations quand on utilise les données fiscales (7 %, contre 18 % dans le cas des

données d'enquête). La première rangée du tableau 20 montre que, dans la population pour laquelle on a utilisé des données d'enquête, 2 207 répondants (15 %) ont dit avoir un emploi rémunéré dans l'interview sur le travail de janvier, mais n'ont pas déclaré de rémunérations dans l'interview sur le revenu. La situation inverse a pu être observée dans 556 cas (4 %).

Les tableaux 22 et 23 présentent une comparaison analogue, mais pour les indemnités des accidents du travail. Parmi les répondants pour lesquels on a utilisé des données d'enquête, 140 ont déclaré avoir touché des indemnités pour accidents du travail dans les interviews sur le travail et le revenu; 178 autres ont déclaré avoir touché des indemnités dans une enquête, mais pas dans l'autre. Ainsi, 2 % des répondants pour lesquels on a utilisé des données d'enquête ont apparemment touché des indemnités pour accidents du travail.

Cette proportion s'établit à 4 % chez les répondants qui ont donné accès à leur dossier fiscal. Cela comprend 305 cas dans lesquels les données de l'interview sur le travail et les données fiscales concordent, 212 cas dans lesquels des répondants ont déclaré des indemnités pour accidents du travail dans leur déclaration d'impôt, mais non dans l'interview sur le travail, et 43 cas dans lesquels la situation inverse était vraie.

Tableau 22. Répondants ayant déclaré des indemnités pour accidents du travail
— Données d'enquête

Code-réponse janv./mai	Total	Pas d'indicateur en janv. et pas de montant en mai	Indicateur en janv., mais pas de montant en mai	Non- réponse en janv.	Pas d'indicateur en janv., mais montant en mai	Indicateur en janv. et montant en mai	Autre
Total	15 055	13 509	115	1 216	63	140	12
Comp/comp	9 446	9 226	25*	2	62**	131	0
Comp/part	842	806	26	0	1**	9	0
Part/comp	103	103	0	0	0	0	0
Part/part	1	1	0	0	0	0	0
Comp/non	3 311	3 246	64	1	0	0	0
Part/non	31	30	0	1	0	0	0
Non/comp	770	21	0	737	0	0	12
Non/part	62	0	0	62	0	0	0
Non/non	489	76	0	413	0	0	0

Tableau 23. Répondants ayant déclaré des indemnités pour accidents du travail
— Données fiscales

Code-réponse janv./mai	Total	Pas d'indicateur en janv. et pas de montant en mai	Indicateur en janv., mais pas de montant en mai	Non- réponse en janv.	Pas d'indicateur en janv., mais montant en mai	Indicateur en janv. et montant en mai	Autre
Total	14 627	12 923	43	1 106	212	305	38
Comp/comp	13 318	12 755	42	5	211**	305	0
Comp/part	125	124	1	0	0	0	0
Part/comp	14	14	0	0	0	0	0
Part/part	1	1	0	0	0	0	0
Comp/non	0	0	0	0	0	0	0
Part/non	0	0	0	0	0	0	0
Non/comp	1 161	29	0	1 093	1	0	38
Non/part	8	0	0	8	0	0	0
Non/non	0	0	0	0	0	0	0

* Imputer le montant

** Changer l'indicateur

Tableau 24. Répondants ayant déclaré des prestations d'assurance-chômage
— Données d'enquête

Code-réponse janv./mai	Total	Pas d'indicateur en janv. et pas de montant en mai	Indicateur en janv., mais pas de montant en mai	Non- réponse en janv.	Pas d'indicateur en janv., mais montant en mai	Indicateur en janv. et montant en mai	Autre
Total	15 055	11 893	683	1 152	151	1 100	76
Comp/comp	9 446	8 147	105*	2	143**	1 049	0
Comp/part	842	648	137	0	6**	51	0
Part/comp	103	103	0	0	0	0	0
Part/part	1	1	0	0	0	0	0
Comp/non	3 311	2 869	441	1	0	0	0
Part/non	31	30	0	1	0	0	0
Non/comp	770	19	0	677	2**	0	72
Non/part	62	0	0	58	0	0	4
Non/non	489	76	0	413	0	0	0

Tableau 25. Répondants ayant déclaré des prestations d'assurance-chômage
— Données fiscales

Code-réponse janv./mai	Total	Pas d'indicateur en janv. et pas de montant en mai	Indicateur en janv., mais pas de montant en mai	Non- réponse en janv.	Pas d'indicateur en janv., mais montant en mai	Indicateur en janv. et montant en mai	Autre
Total	14 627	10 755	54	889	402	2 273	254
Comp/comp	13 318	10 604	53	5	391**	2 265	0
Comp/part	125	115	1	0	1**	8	0
Part/comp	14	13	0	0	1**	0	0
Part/part	1	1	0	0	0	0	0
Comp/non	0	0	0	0	0	0	0
Part/non	0	0	0	0	0	0	0
Non/comp	1 161	22	0	876	9**	0	254
Non/part	8	0	0	8	0	0	0
Non/non	0	0	0	0	0	0	0

* Imputer le montant

** Changer l'indicateur

Dans le cas de l'assurance-chômage, 1 100 répondants (7 %) pour lesquels on a utilisé des données d'enquête ont déclaré avoir reçu des prestations d'assurance-chômage dans les deux interviews, 683 (5 %) ont déclaré en avoir reçu dans l'interview sur le travail, mais non dans l'interview sur le revenu, et 151 (1 %) ont fait l'inverse (tableau 24). Chez les répondants qui ont donné accès à leur dossier fiscal, des écarts ont été observés dans 3 % des cas (tableau 25). Le nombre de ceux qui ont déclaré avoir reçu des prestations

d'assurance-chômage dans l'enquête sur le travail et dans leur déclaration d'impôt s'établissait à 2 273 (16 %).

5. MÉTHODES D'IMPUTATION ET DE CALCUL

L'imputation de certains éléments de revenu manquants dans les données de la première phase a été faite à l'aide du Système généralisé de vérification et d'imputation (SGVI). Les valeurs qui pouvaient être facilement calculées en fonction de paramètres connus ont fait l'objet d'une imputation déterministe. En gros, la stratégie retenue peut s'énoncer ainsi :

- *Imputation en fonction du plus proche voisin pour la majorité des variables sur le revenu* : dans le SGVI, le même voisin sert à imputer la totalité des variables manquantes, ce qui permet de sauvegarder les relations entre les divers éléments.
- *Imputation déterministe des variables suivantes* : prestation fiscale pour enfants, crédit pour la taxe sur les produits et services, pension de sécurité de la vieillesse, supplément de revenu garanti et allocation de conjoint.
- Une *régression linéaire* a servi à établir un modèle de l'impôt fédéral et provincial exigible des répondants qui n'ont pas donné accès à leur dossier fiscal.

5.1 Système généralisé de vérification et d'imputation (SGVI)

Le Système généralisé de vérification et d'imputation (SGVI) comprend une vérification, une localisation d'erreurs et une imputation. La vérification permet de voir si certaines valeurs d'un dossier sont incorrectes, manquantes, incompatibles ou exceptionnelles.

L'imputation permet de remplacer une valeur erronée par une valeur plus plausible. Ces deux fonctions sont unies par la fonction de localisation d'erreurs, qui repère les champs d'un dossier qui doivent faire l'objet d'une imputation.

5.2 Imputation en fonction du plus proche voisin à l'aide du SGVI

L'imputation en fonction du plus proche voisin se fait en deux étapes : l'imputation partielle et l'imputation totale. Dans l'imputation partielle, un indicateur est affecté à un certain nombre de champs pour lesquels nous posons que le répondant a touché un revenu de la source en question et pour lesquels il nous faut un donneur ayant déclaré un revenu de la même source. Dans l'imputation d'une non-réponse totale, par contre, les sources de revenu qui devraient être imputées ne sont pas connues. Nous disposons par conséquent de moins de renseignements pour faire l'imputation, et nous devons imputer à la fois des sources et des montants en fonction du donneur. Ainsi, si le donneur a touché des revenus de placements et des revenus de pension, ces revenus sont attribués au non-répondant.

Les répondants de 16 ans et plus admissibles à l'interview sur le revenu ont été répartis en fonction de règles précises en quatre groupes : donneurs, receveurs (non-réponse partielle), receveurs (non-réponse totale) et répondants exclus de l'imputation. Cette quatrième catégorie comprend les dossiers dont les données ne nécessitaient aucune intervention, mais dont nous avons des raisons de croire qu'ils pouvaient contenir des incompatibilités. Ainsi, seuls des dossiers sans réserve ont été utilisés dans l'imputation.

Tableau 26. Statut d'imputation de répondants admissibles à l'interview sur le revenu

	Donneur	Receveur		Exclus de l'imputation	Total
		Non-réponse partielle	Non-réponse totale		
Données d'enquête	8 046 39,8 %	1 661	117	1 390	11 214
Données de Revenu Canada	12 160 60,2 %	0	0	2 467	14 627
Aucune donnée	0	0	3 829	0	3 829
Total	20 206 100 %	1 661	3 946	3 857	29 670

5.2.1 Imputation : «receveurs — non-réponse partielle»

Dans cette étape, nous avons imputé des montants pour des répondants ayant donné des réponses partielles. Pour choisir des donneurs qui conviennent, nous avons divisé le bassin des donneurs en 8 groupes d'appariement :

Hommes 16-24	Femmes 16-14
Hommes 25-54	Femmes 25-54
Hommes 55-69	Femmes 55-69
Hommes 70+	Femmes 70+

Les variables d'appariement discrètes (les *groupes de vérification* du SGVI) étaient les suivantes :

Région (Atlantique, Québec, Ontario, Prairies et Colombie-Britannique)

État matrimonial (marié ou union libre, divorcé ou séparé, veuf et célibataire)

Région urbain ou région rurale

Niveau d'instruction (moins qu'un niveau secondaire, secondaire à moins qu'un niveau universitaire, grade universitaire ou niveau supérieur)

Étudiant à plein temps ou à temps partiel

Nombre d'enfants de 0 à 15 ans

Nombre d'enfants de 16 à 18 ans

Nombre d'adultes

A touché des prestations d'assurance-chômage

A touché des indemnités pour accidents du travail

A touché des prestations d'assistance sociale

RPC/RRQ (valeur prévue)

Indicateur de travailleur rémunéré

Indicateur de travailleur agricole indépendant (constitué en corporation)

Indicateur de travailleur agricole indépendant (non constitué en corporation)

Indicateur de travailleur non rémunéré

Emploi ou non

Emploi agricole ou non

Conjoint ayant touché des prestations de bien-être ou non

Il convient de noter que les variables qui concernent l'élément «travail» ne sont pas disponibles dans le cas des deux groupes de 70 ans et plus.

Les variables d'appariement continues (*appariement obligatoire* dans le SGVI) étaient les suivantes :

Nombre d'emplois occupés pendant l'année

Nombre de semaines de chômage

Nombre habituel d'heures travaillées pour tous les emplois

Rémunérations annuelles totales

Ensemble des variables du revenu ne devant pas à être imputées

Il convient de noter que les variables qui concernent l'élément «travail» ne sont pas disponibles dans le cas des deux groupes de 70 ans et plus. L'imputation avec le SGVI terminée, il restait 63 cas qui ont dû faire l'objet d'une imputation manuelle.

5.2.2 Imputation : «receveurs — non-réponse totale»

Dans le cas de ces receveurs, les variables du revenu devaient toutes être imputées. Les groupes d'appariement et les variables d'appariement discrètes étaient les mêmes que ceux qui avaient été utilisés dans l'imputation partielle. Toutefois, seules les variables d'appariement continues suivantes ont été utilisées :

Nombre d'emplois occupés pendant l'année

Nombre de semaines de chômage

Nombre habituel d'heures travaillées pour tous les emplois

Rémunérations annuelles totales

Revenu total de l'interview sur le revenu (quand ce renseignement était disponible)

L'imputation avec le SGVI terminée, il restait 94 cas qui ont dû faire l'objet d'une imputation manuelle.

5.2.3 Imputation déterministe

Dans le cas des répondants pour lesquels on a utilisé les données d'enquête, une imputation déterministe a été faite pour les variables suivantes : prestation fiscale pour enfants, PSV/SRG/AC et crédit pour la TPS. Dans le cas des répondants qui ont donné accès à leur dossier fiscal, l'imputation déterministe a été utilisée pour les crédits de TPS.

L'imputation déterministe a été faite après l'imputation fondée sur la méthode du plus proche voisin. Les valeurs ont été calculées en fonction des caractéristiques des répondants et des paramètres des programmes. Quand cela était justifié, nous avons comparé la valeur calculée au montant déclaré par le répondant. Si l'écart entre les deux montants était trop important (à en juger par certains critères), nous avons utilisé le montant calculé; sinon, nous avons gardé le montant déclaré, afin de nous en tenir autant que possible aux renseignements donnés par les répondants.

Nous avons calculé les crédits pour la TPS des répondants dont les données fiscales ont été utilisées parce que ce montant ne figure pas dans la déclaration d'impôt. Les allocations familiales des répondants du Québec ont été calculées pour la même raison.

5.2.4 Imputation de l'impôt fédéral et provincial sur le revenu

Un modèle de l'impôt fédéral et provincial exigible a été établi par régression linéaire, une fois l'imputation déterministe effectuée. L'impôt fédéral et provincial exigible a été calculé pour tous les dossiers, sauf quand les données venaient de fichiers d'impôt. Dans le cas du Québec, l'impôt provincial sur le revenu a été calculé pour tous les dossiers, puisque l'impôt provincial sur le revenu pour le Québec ne figure pas dans le fichier d'impôt.

5.3 Évaluation de l'imputation en fonction du plus proche voisin à l'aide du SGVI

Pour apprécier la qualité de l'imputation, on peut par exemple se demander combien de fois le même donneur a été choisi. Le tableau 27 montre que 81 % des dossiers utilisés comme donneurs n'ont été choisis qu'une seule fois. On y apprend également que la majorité des donneurs (59 %) étaient des répondants qui avaient donné accès à leur dossier fiscal. Cette proportion se compare à la composition du bassin des donneurs : comme le montre le tableau 27, plus de 12 000 des quelque 20 000 donneurs, soit 60 %, étaient des répondants qui avaient donné accès à leur dossier fiscal. Une faible minorité de donneurs ont été utilisés plusieurs fois, et ce fait tient à ce que le bassin des donneurs était passablement petit par comparaison au nombre de cas nécessitant une imputation. Cette situation devrait s'améliorer pour l'année de référence 1996, avec la création du deuxième panel, le bassin de donneurs se trouvant doublé.

Tableau 27. Source de donneurs (données d'enquête et données fiscales) et nombre de fois que certains donneurs ont été choisis

Source des données	Nombre de fois qu'un même donneur a été choisi						Total
	1	2	3	4	5	6+	
Enquête	1 425 <i>79,5 %</i>	265 <i>14,8 %</i>	60 <i>3,3 %</i>	27 <i>1,5 %</i>	8 <i>0,4 %</i>	7 <i>0,4 %</i>	1 792 41 %
Impôt	2 096 <i>81,3 %</i>	383 <i>14,9 %</i>	65 <i>2,5 %</i>	17 <i>0,7 %</i>	7 <i>0,3 %</i>	10 <i>0,4 %</i>	2 578 59 %
Total	3 521 80,6 %	648 14,8 %	125 2,9 %	44 1 %	15 0,3 %	17 0,4 %	4 370 100 %

Diverses autres mesures ont été prises pour évaluer la qualité de l'imputation. On a notamment comparé des moyennes, des médianes et des intervalles de la population ayant fait l'objet d'imputations par comparaison aux données fiscales de la même population.

Évaluation de l'imputation totale

Chez les personnes de 16 à 69 ans, la comparaison des données fiscales et des données imputées a été fondée sur les 2 656 observations qui correspondaient au dossier fiscal (tableau 28).

Le tableau appelle quelques explications. Chaque variable de revenu du tableau comporte une ligne qui présente les résultats de l'imputation et une ligne qui contient les données vraies (dossier fiscal). La quatrième colonne contient le nombre de dossiers dans lesquels le montant imputé n'était pas zéro. Les deux colonnes suivantes présentent respectivement la moyenne et la médiane, exclusion faite des dossiers où la valeur imputée et la valeur vraie étaient toutes deux égales à zéro.

Si l'on excepte quelques variables, l'imputation semble avoir donné des résultats très raisonnables. Les variables où les résultats se sont pas tout à fait ceux qu'on attendrait sont celles pour lesquelles nous avons eu de la difficulté à trouver un donneur qui convenait. Par exemple, il est difficile d'imputer des valeurs négatives importantes pour un revenu d'un travail indépendant agricole, et cela explique que le montant moyen imputé est passablement plus élevé que la moyenne tirée des données fiscales. Des chiffres analogues à ceux du tableau 28 ont été calculés par groupe d'âge quinquennal et pour les personnes de 70 ans et plus.

Tableau 28. Comparaison de la moyenne et de la médiane de certains éléments de revenu, dossiers imputés et dossiers fiscaux correspondants : imputation totale, personnes de 16 à 69 ans

	Valeur	N	Montant ≠ 0	Moyenne (\$)	Médiane (\$)	Intervalle
Rémunérations	impôt	2 600	1 923	23 800	17 100	0 à 1 460 000
	imputée	2 600	1 922	23 300	20 200	0 à 175 000
Travail indépendant (agricole)	impôt	126	114	4 800	1 300	-25 000 à 60 000
	imputée	126	62	7 400	0	-9 000 à 60 000
Travail indépendant (non agricole)	impôt	373	307	11 600	2 500	-65 000 à 230 000
	imputée	373	206	10 600	0	-90 000 à 300 000
Placements	impôt	1 446	997	2 200	100	-85 000 à 155 000
	imputée	1 446	889	1 600	200	-30 000 à 310 000
Gains en capital	impôt	228	176	5 400	100	0 à 180 000
	imputée	228	124	6 200	100	0 à 260 000
RPC/RRQ	impôt	323	307	4 600	0	0 à 15 000
	imputée	323	210	3 700	4 700	0 à 45 000
A.-C.	impôt	608	587	5 000	3 200	0 à 21 000
	imputée	608	432	3 700	2 400	0 à 19 000
Assistance sociale	impôt	204	168	4 700	3 300	0 à 21 000
	imputée	204	120	3 300	1 400	0 à 18 000
Indemnités pour accidents du travail	impôt	110	106	4 300	1 600	0 à 33 000
	imputée	110	54	2 100	0	0 à 42 000
Pension d'employeur	impôt	286	187	7 600	2 800	0 à 76 000
	imputée	286	177	7 900	3 300	0 à 57 000
Pension alimentaire	impôt	80	41	3 800	300	0 à 59 000
	imputée	80	45	2 700	1 400	0 à 18 000
Autres revenus imposables	impôt	649	448	2 500	400	0 à 77 000
	imputée	649	272	1 400	0	0 à 62 000

Évaluation de l'imputation partielle

Les résultats des calculs qui ont été faits pour les dossiers nécessitant une imputation partielle sont présentés au tableau 29, ici encore pour les personnes de 16 à 69 ans. Dans le cas des rémunérations, la moyenne des montants imputés est supérieure à celle des données fiscales, mais pas assez pour que cela soit inquiétant. On observe en outre que le montant maximum des données fiscales est supérieur à celui des données imputées. Comme nous l'avons déjà souligné, le revenu d'un travail indépendant agricole est difficile à imputer quand les montants sont peu élevés ou négatifs; la moyenne des valeurs imputées est finalement plus élevée que celle des données fiscales. Dans le cas du revenu d'un travail indépendant non agricole, la moyenne ainsi que les montants minimum et maximum imputés sont tous inférieurs à ceux des données fiscales. Les revenus de placements ont été légèrement sous-estimés, alors que la moyenne des gains en capital a été surestimée, principalement en raison de l'imputation d'un petit nombre de montants importants. À l'inverse, la surestimation des rentes du RPC/RRQ a probablement été causée par l'indicateur fondé sur les renseignements de l'EPA, et au sujet duquel nous avons déjà donné des explications. En ce qui concerne l'assistance sociale, il est difficile d'évaluer les résultats obtenus sur le plan individuel, puisque cette variable devrait être étudiée au niveau de la famille. Les résultats obtenus à l'égard des autres variables ont été satisfaisants.

Des renseignements complémentaires sur l'évaluation de l'imputation partielle peuvent être obtenus sur demande.

Tableau 29. Comparaison de la moyenne et de la médiane de certains éléments de revenu, dossiers imputés et dossiers fiscaux correspondants : imputation partielle, personnes de 16 à 69 ans

	Valeur	N	Montant≠0	Moyenne (\$)	Médiane (\$)	Intervalle
Rémunérations	impôt	268	113	10 500	4 600	0 à 86 000
	imputée	268	139	14 100	9 200	0 à 52 600
Travail indépendant (agricole)	impôt	478	25	-300	-300	-11 000 à 20 000
	imputée	478	13	-1 900	0	-7 000 à 26 600
Travail indépendant (non agricole)	impôt	465	35	2 600	0	-11 200 à 36 600
	imputée	465	30	800	0	-33 300 à 18 200
Placements	impôt	478	155	1 700	100	-4 200 à 59 000
	imputée	478	150	1 200	100	- 2 700 à 33 700
Gains en capital	impôt	492	34	7 500	0	0 à 145 100
	imputée	492	28	14 400	50	0 à 260 000
RPC/RRQ	impôt	101	56	2 900	500	0 à 9 610
	imputée	101	101	4 800	5 000	100 à 11 234
A.-C.	impôt	222	158	4 500	3 500	0 à 16 500
	imputée	222	173	4 800	3 700	0 à 19 000
Assistance sociale	impôt	107	34	2 900	800	0 à 18 000
	imputée	107	62	4 300	2 600	0 à 19 500
Indemnités pour accidents du travail	impôt	88	22	3 500	700	0 à 31 000
	imputée	88	37	4 100	1 400	0 à 24 000
Pension d'employeur	impôt	478	28	5 600	300	0 à 108 600
	imputée	478	38	6 600	4 300	0 à 30 100
Pension alimentaire	impôt	479	5	2 000	100	0 à 8 200
	imputée	479	5	2 000	100	0 à 9 400
Autres revenus imposables	impôt	485	92	1 600	200	0 à 32 100
	imputée	485	61	900	0	0 à 30 800

5.4 Revenu total

Après avoir fait l'imputation des non-réponses totales et partielles à l'aide du SGVI, il restait à calculer le revenu total final. Cet aspect de l'exploitation des données comportait plusieurs étapes. Pour commencer, avant de calculer le revenu total, il a fallu faire une imputation déterministe de certains éléments du revenu total (prestation fiscale pour enfants, pension de sécurité de la vieillesse/supplément de revenu garanti et crédit pour la taxe sur les produits et services). Pour faire chacune de ces imputations, nous avons dû calculer un revenu total pour cette étape en particulier. Le tableau 30 ci-dessous présente succinctement les divers totaux ainsi que les étapes auxquelles ils se rapportent.

Le premier revenu total créé a été le revenu total nécessaire au calcul de la pension de sécurité de la vieillesse/du supplément de revenu garanti/de l'allocation de conjoint. Cela a permis de calculer l'élément PSV/SRG des personnes qui y étaient admissibles et qui auraient dû en recevoir (en fonction des critères d'admissibilité du programme), mais qui n'en ont pas déclaré. Une fois que le revenu de PSV/SRG a été ainsi attribué (uniquement aux répondants admissibles au sujet desquels on a utilisé des données d'enquête), deux autres revenus totaux nécessaires au calcul de la TPS/PFE et de l'impôt fédéral/provincial ont été calculés (la PSV entrant dans le calcul de chacun de ces revenus totaux). Ces deux totaux ont été utilisés dans le calcul de ces éléments. La PFE et l'impôt fédéral/provincial ont été calculés uniquement pour les répondants admissibles au sujet desquels on a utilisé des données d'enquête, alors que la TPS a été calculée pour l'ensemble des répondants, puisqu'il n'y a pas de données fiscales sur cet élément. L'imputation des éléments de revenu se trouvant ainsi terminée, il nous était possible de calculer le revenu monétaire total.

Tableau 30. Sommaire des revenus totaux de l'EDTR

Revenus totaux	Exclusions
Revenu total utilisé dans le calcul de l'admissibilité à la pension de sécurité de vieillesse/au supplément de revenu garanti/à l'allocation de conjoint	Revenus non imposables comme la prestation fiscale pour enfants, le crédit pour la TPS et les prestations d'assistance sociale
Revenu total au sens de l'impôt sur le revenu utilisé dans le calcul de l'admissibilité au crédit pour la TPS et à la prestation fiscale pour enfants	Éléments comme la prestation fiscale pour enfants et le crédit pour la TPS, qui ne figurent pas dans la déclaration d'impôt (équivalent de la ligne 150 de la déclaration d'impôt)
Revenu total utilisé dans le calcul de l'impôt fédéral et provincial	Revenus non imposables comme la prestation fiscale pour enfants, le crédit pour la TPS, les prestations d'assistance sociale, le supplément de revenu garanti/l'allocation de conjoint (comprend cependant la pension de sécurité de la vieillesse)
Revenu monétaire total	Revenus non monétaires comme les gains en capital et les retraits de REER

6. REVENU FAMILIAL ET FAIBLE REVENU

Après avoir calculé les revenus individuels (éléments de revenu et revenu total), il restait à calculer les variables du revenu familial. Dans l'EDTR, la «famille» est la famille économique, c'est-à-dire l'ensemble des personnes apparentées par le sang, par alliance, par adoption ou par union de fait qui vivaient ensemble dans le même logement le 31 décembre 1993. Il convient de noter qu'une personne n'avait pas à faire partie de la famille toute l'année; il suffisait qu'elle se trouve dans la famille à la fin de l'année. Ainsi,

chaque personne s'est vue attribuer un revenu familial représentant la somme des revenus annuels de chacun des membres de la famille.

En ce qui concerne le faible revenu, un seuil de faible revenu (SFR) a été attribué à chaque personne en fonction de la taille de sa famille et de la taille de sa collectivité. Les SFR étaient identiques à ceux de l'Enquête sur les finances des consommateurs et ils s'appuyaient sur des données de l'Enquête sur les dépenses des familles de 1992. Les dépenses moyennes des familles au titre de l'alimentation, du logement et de l'habillement servent de base à l'établissement des SFR. Le seuil correspond au niveau de revenu où la famille moyenne d'une taille donnée et d'une collectivité d'un type donné consacre au moins 54,7 % de son revenu à ces biens de première nécessité.

Le SFR des personnes seules et celui des familles de 2 à 7 personnes et plus diffèrent. De plus, le SFR varie en fonction du lieu de résidence : région rurale, petite agglomération urbaine (moins de 30 000 habitants), 30 000 à 99 000, 100 000 à 499 999 et 500 000 et plus. Il s'ensuit qu'il y a 35 SFR. Le fichier contient le SFR qui correspond à la taille de la famille et à la taille de la collectivité du répondant. Les membres d'une même famille ont tous le même SFR.

L'EDTR et l'EFC diffèrent quelque peu en ce qui concerne le calcul de la variable *taille de la région de résidence* (taille de la collectivité); ces différences ont eu un certain effet sur les estimations des faibles revenus. La méthode de calcul sera modifiée pour les données de la deuxième phase et elle sera décrite dans un autre document de travail sur la dynamique du travail et du revenu.