

**N° 97-04 au catalogue**

**SÉLECTION D'UN ÉCHANTILLON SUPPLÉMENTAIRE  
POUR LA PRODUCTION D'ESTIMATIONS  
TRANSVERSALES SUR LE REVENU**

Numéro d'enregistrement du produit 75F0002M

Janvier 1997

Michel Latouche, Division des méthodes d'enquêtes sociales

Sylvie Michaud, Division des méthodes d'enquêtes sociales

Johanne Tremblay, Division des méthodes d'enquêtes-ménages

Ruth Dibbs, Division d'enquêtes des ménages

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<sup>e</sup> étage, section C6, Statistique Canada, Ottawa (Ontario), Canada, K1A 0T6; par INTERNET: [DYNAMIQUE@STATCAN.CA](mailto:DYNAMIQUE@STATCAN.CA); par téléphone au (613) 951-4633; ou par télécopieur au (613) 951-3012.



## SOMMAIRE

L'intégration de la collecte de données sur le revenu pour l'EDTR et l'EFC est prévue pour l'année fiscale 1997. Les plans actuels prévoient que la majorité des données proviendront de l'échantillon de l'EDTR. En effet, pour l'année de référence de 1997, deux panels de l'EDTR seront actifs. Cependant, le premier panel en sera à sa cinquième vague de collecte et le second panel à sa deuxième vague (de collecte). Le rapport présente un certain nombre d'analyses sommaires qui justifient les recommandations pour la sélection de l'échantillon supplémentaire. Nous avons tout d'abord essayé d'évaluer l'impact de l'intégration sur la taille de l'échantillon transversale, sur la variabilité des estimations et sur la représentativité de l'échantillon. Puis nous avons étudié différentes options comme un échantillon de logements ou un échantillon de ménages, le moment approprié pour sélectionner l'échantillon et le ou les groupe(s) de rotation de l'EPA à réserver. *À court terme, afin d'améliorer la représentativité transversale, il est recommandé d'ajouter un échantillon supplémentaire de 8,725 ménages correspondant à un groupe de rotation entier de l'EPA. Considérant qu'un échantillon de logements assurerait une meilleure représentativité transversale de l'échantillon mais que les coûts associés seraient beaucoup plus élevés et entraîneraient une diminution considérable de la taille de l'échantillon, il est recommandé qu'un échantillon de ménages soit choisi en août 1997. Le groupe de rotation 4 se trouvant dans l'échantillon de l'EPA de avril 1997 à septembre 1997 devrait être réservé pour l'échantillon supplémentaire.*

La solution proposée est cependant une solution à court terme. En effet, le manque de temps et d'information ne permet pas d'évaluer adéquatement les impacts complets de la stratégie proposée. Diverses études se feront tout au cours du

projet d'intégration pour tenter de quantifier les impacts dans la mesure du possible. Il se peut qu'il y ait une détérioration de la qualité des estimations produites, s'il s'avérait qu'il existe un certain biais de réponse dû à l'érosion et que les ajustements de non-réponse ne compensent pas adéquatement. La stratégie proposée devrait donc être réévaluée lors des deux années suivant l'intégration.

## TABLE DES MATIÈRES

	Page
1. Introduction	1
2. Intégration et impact sur la taille de l'échantillon et sur la variabilité	2
3. Intégration et impact sur la représentativité transversale de l'échantillon	6
4. Allocation de l'échantillon supplémentaire	8
5. Date de sélection de l'échantillon et unité d'échantillonnage logement vs ménage	10
6. Choix de l'échantillon (groupe de rotation)	11
7. Besoins à moyen terme	13
Annexe 1 : Caractéristiques des personnes selon la réponse aux interviews: Panel complet de l'enquête SIPP de 1984	15
Annexe 2 : Définition des RMR	18
Annexe 3 : Taux de réponse longitudinal sur trois ans de l'EDTR	19
Annexe 4 : Sélection de l'échantillon supplémentaire supposant le même contenu	24
Annexe 5 : Options pour l'échantillon supplémentaire de 1998	26
Annexe 6 : Choix de l'échantillon	31



## 1. Introduction

L'intégration de la collecte de données sur le revenu pour l'EDTR et l'EFC est prévue pour l'année fiscale 1997. Les plans actuels prévoient que la majorité des données proviendront de l'échantillon de l'EDTR. En effet, pour l'année de référence de 1997, deux panels de l'EDTR seront actifs. Cependant, le premier panel en sera à sa cinquième vague de collecte et le second panel à sa deuxième vague (de collecte). Le rapport présente un certain nombre d'analyses sommaires qui justifient les recommandations pour la sélection de l'échantillon supplémentaire. Nous avons tout d'abord essayé d'évaluer l'impact de l'intégration sur la taille de l'échantillon transversale, sur la variabilité des estimations et sur la représentativité de l'échantillon. Puis nous avons étudié différentes options comme un échantillon de logements ou un échantillon de ménages, le moment approprié pour sélectionner l'échantillon et le ou les groupe(s) de rotation de l'EPA à réserver. *À court terme, afin d'améliorer la représentativité transversale, il est recommandé d'ajouter un échantillon supplémentaire de 8,725 ménages correspondant à un groupe de rotation entier de l'EPA. Considérant qu'un échantillon de logements assurerait une meilleure représentativité transversale de l'échantillon mais que les coûts associés seraient beaucoup plus élevés et entraîneraient une diminution considérable de la taille de l'échantillon, il est recommandé qu'un échantillon de ménages soit choisi en août 1997. Le groupe de rotation 4 se trouvant dans l'échantillon de l'EPA de avril 1997 à septembre 1997 devrait être réservé pour l'échantillon supplémentaire.*

La solution proposée est cependant une solution à court terme. En effet, le manque de temps et d'information ne permet pas d'évaluer adéquatement les impacts complets de la stratégie proposée. Diverses études se feront tout au cours du projet d'intégration pour tenter de quantifier les impacts dans la mesure du

possible. Il se peut qu'il y ait une détérioration de la qualité des estimations produites, s'il s'avérait qu'il existe un certain biais de réponse dû à l'érosion et que les ajustements de non-réponse ne compensent pas adéquatement. La stratégie proposée devrait donc être réévaluée lors des deux années suivant l'intégration.

Tout au long du rapport, on parle d'année de référence. L'année de référence 1993 réfère au revenu associé à l'année de calendrier 1993. La collecte de ces données sur le revenu se fait en mai de l'année suivante. La section 1 discute de l'impact de l'intégration sur la taille effective de l'échantillon transversale et sur la variabilité des estimations. La section 2 discute de l'impact de l'intégration sur la représentativité de l'échantillon transversale. La section 3 propose différents scénarios d'allocation d'un échantillon supplémentaire. La section 4 compare un échantillon de logement et un échantillon de ménages en terme de représentativité et de coûts. La section 5 discute du choix du groupe de rotation de l'échantillon de l'EPA qui composera l'échantillon supplémentaire.

## **2. INTÉGRATION ET IMPACT SUR LA TAILLE DE L'ÉCHANTILLON ET SUR LA VARIABILITÉ**

L'intégration des deux enquêtes aura un impact sur les estimations de revenu, d'un effet différent selon les niveaux d'estimation et/ou les variables étudiées. En effet, la conception des enquêtes, une de nature purement transversale et l'autre à vocation principalement longitudinale, laisse supposer des impacts. Une étude comparative a été effectuée sur la première vague de données de l'EDTR mais le manque de données longitudinales et l'absence du second panel limite les comparaisons. Nous avons fait l'hypothèse que l'enquête intégrée doit maintenir les niveaux de précision actuels de l'EFC.

La première partie de l'étude s'est donc penchée sur la comparaison des tailles d'échantillon des deux enquêtes. Le premier panel de l'EDTR a été introduit pour l'année de référence 1993. À la première vague de l'entrevue, environ 15000 ménages ont été interviewés. Comme l'EDTR suit les individus pendant six ans, on observe une érosion de l'échantillon (i.e. une non-réponse cumulée) à cause par exemple de l'impossibilité de dépister un pourcentage de l'échantillon de gens qui déménagent. Ce fait est en partie balancé par le fait que l'échantillon est dynamique. Parce que la composition d'un ménage est dynamique dans le temps, un certain nombre de personnes vivant initialement ensemble quittent le ménage original pour créer des nouveaux ménages. De nouveaux individus viennent aussi s'installer dans des ménages existants. À chaque année, les individus qui restent avec les répondants longitudinaux et qui font partie de ces nouveaux ménages sont interviewés. Le tableau 1a décrit le nombre de ménages interviewés pour le premier panel de l'EDTR. On constate une baisse marquée du nombre de répondants au début du panel, suivie d'une stagnation causée par un équilibre entre les ménages nonrépondants et les ménages créés. D'autres analyses (non présentées dans ce document) ont démontrées un phénomène similaire au niveau provincial.

**Tableau 1a.** Nombre de ménages répondants au premier panel.

	Préliminaire*	vague 1	vague 2	vague 3	vague 4**	vague 5**	vague 6**
Ménages répondants	15,006	14,694	15,224	15,507	15,434	15,445	15,456
% de l'échantillon préliminaire	88	86	91	92	91	91	91

\* En fait un échantillon de 20,000 ménages avait été sélectionné, plus de 17,000 de ces ménages ont répondu mais des réductions de budget ont limité la taille de l'échantillon à 15,000 ménages

Un deuxième panel sera introduit pour l'EDTR pour l'année de référence 1996. Ce nouveau panel est formé de deux groupes de rotation de l'échantillon remanié de l'Enquête sur la Population Active (EPA). Pour l'année de référence 1997, le second panel devrait être à sa deuxième année de collecte et le premier panel à sa cinquième année. On peut tenter de faire une inférence sur le nombre de ménages éligibles à la collecte pour l'année de référence 1997 en faisant des prédictions à l'aide des taux de croissance/décroissance du premier panel.

**Tableau 1b.** Projection du nombre de ménages du second panel.

	Prél.+ vague 1	vague 2	vague 3	vague 4	vague 5	vague 6
Ménages répondants	14,229	15,056	15,221	15,056	15,056	15,056
% de l'échantillon préliminaire*	86	91	92	91	91	91

\* taux du premier panel.

Pour la première année de l'intégration, la taille du panel 1 est estimée à 15,445 répondant, celui du panel 2 à 15,056 pour un total de 30,501 ménages. Ces estimations sont calculées à partir de taux de réponse qui incluent des ménages pour lesquels les données sur le revenu ont été imputés. Environ 12% des données sur le revenu sont imputées (imputation totale). Donc pour rendre les taux de réponse comparables à ceux de l'EFC, le taux de l'EDTR devrait être diminué. On pourrait donc s'attendre à avoir un échantillon d'environ 27,000 ménages répondants.

L'EFC choisit son échantillon parmi les répondants à l'EPA. Quatre groupes de rotation sont sélectionnés pour l'EFC. Ça correspond donc à un échantillon de 34,900 ménages. De façon traditionnelle, l'EFC a un taux de réponse d'environ 80% ce qui correspondrait à un échantillon d'environ 27,920 répondants.

L'échantillon de l'EDTR semble donc être légèrement inférieur à celui de l'EFC en terme de taille. Donc si on se fiait seulement sur la taille de l'échantillon pour fins de comparaison, on pourrait supposer que l'échantillon de l'EDTR a besoin d'être augmenté d'environ 1000 ménages.

Les coefficients de variation pour l'estimation du revenu personnel total avant impôt et pour le nombre d'individus de 15 ans et plus avec un revenu familial avant impôt sous le LICO sont présentés pour l'EDTR et l'EFC dans le tableau 2. Les CV réfèrent à l'année de référence 1993. Au niveau canadien, les estimations de l'EDTR restent bonnes pour le revenu personnel considérant que l'EDTR utilise seulement deux groupes de rotation. Pour le nombre d'individus sous le LICO, les différences au niveau canadien sont plus importantes, l'EDTR affichant un CV presque deux fois plus élevé que l'EFC. Au niveau provincial, les différences varient beaucoup selon la province et la variable d'intérêt. Dans certains cas les CV de l'EDTR sont très proches de ceux de l'EFC même avec un échantillon plus petit. Dans d'autres cas les différences sont élevées même si on considère que l'EFC a une taille d'échantillon correspondant au double de celle de l'EDTR.

**Tableau 2.** Comparaison des CV de l'EDTR et de l'EFC pour le revenu personnel avant impôt et pour le nombre d'individus de 15 ans et plus avec un revenu familial avant impôt sous le LICO

PROVINCE	Revenu personnel avant impôt		Nombre d'individus sous le LICO	
	EDTR (2 "rotations")	EFC (4 "rotations")	EDTR (2 "rotations")	EFC (4 "rotations")
TN	7.2	1.9	16.4	6.3
IPE	1.6	2.3	15.6	12.2
NE	3.7	3.2	12.2	6.9
NB	7.2	1.4	10.0	5.6
PQ	2.4	1.2	6.4	3.9
ON	1.4	1.1	8.4	3.5
MA	3.0	1.9	6.2	6.3
SK	2.3	1.5	6.8	5.2
ALB	2.0	2.4	6.9	5.2
CB	1.3	2.1	7.6	6.0
<b>Canada</b>	<b>0.9</b>	<b>0.7</b>	<b>3.5</b>	<b>1.9</b>

### 3. INTÉGRATION ET IMPACT SUR LA REPRÉSENTATIVITÉ TRANSVERSALE DE L'ÉCHANTILLON

Même si en l'absence d'un échantillon supplémentaire on obtenait un échantillon de taille semblable, il est raisonnable de se demander si un tel échantillon serait représentatif transversalement. En effet, l'expérience d'autres enquêtes longitudinales suggère qu'il existe des différences entre les individus répondants et les non-répondants dans les enquêtes longitudinales. Bien que les procédures de pondération tentent de compenser pour ces différences, il est toujours difficile d'évaluer l'impact final. De plus, comme l'EDTR combine à la fois des variables

de revenu et des variables de travail, il n'est pas clair que les ajustements seront efficaces de façon égale pour les deux composantes.

L'hypothèse qui mentionne les différences entre les non-répondants d'érosion et les répondants est supportée dans la littérature. L'annexe 1 compare l'échantillon longitudinal de SIPP ("Survey of Income Program Participation" faite aux États-Unis), en décomposant l'échantillon selon diverses composantes; répondants à toutes les entrevues, érosion, manquant une entrevue seulement et manquant deux entrevues et plus. Lorsqu'on compare les distributions, on peut voir que similairement à ce qui est observé au Canada, l'érosion est plus grande dans certaines régions (par exemple, Los Angeles), dans les grandes villes, parmi les gens qui vivaient dans un logement loué à la première vague. Ce sont des résultats similaires à ce que l'EDTR a observé lors de la première vague. Le tableau 3 compare les estimés de famille de taille 1 vs 2+ lors de la première vague de l'EDTR et de l'EFC (un certain nombre de caractéristiques démographiques ont été comparées, mais cette caractéristique semblait être importante).

**Tableau 3.** Distribution du nombre de personnes selon la taille de la famille 1 vs 2+ pour l'EDTR et l'EFC  
(après post-stratification, poids intégré ménage dans les deux cas)

	EFC %	EDTR %	Différence des estimations totales * (EDTR-EFC)/EFC
familles de taille 1	13.4	12.9	- 4.0 %
familles de taille 2+	86.6	87.1	+0.4 %

\*Cette différence est faite à partir des estimations de nombre de personnes. Elle ne peut donc pas être reproduite directement de ce tableau.

L'EDTR semble sous représenter les familles de taille 1 et sur-représenter les familles de taille 2+. Ceci peut probablement être expliqué par le fait que l'échantillon de l'EDTR doit dépister les individus qui déménagent. On perd donc un certain nombre d'individus qui quittent leur ménage original et forment des ménages uniques donc des familles de taille 1. Une bonne répartition entre ces types de familles est très importante pour les estimations produites par l'EFC. Il se peut aussi que les déménageurs non dépistés aient des caractéristiques différentes des gens qui n'ont pas déménagé. Un échantillon supplémentaire permet de compenser en partie pour des biais de représentativité qui pourraient se produire dans l'échantillon à cause de l'érosion. Donc bien que la perte nette en terme de taille d'échantillon soit petite, il est quand même recommandé de sélectionner un échantillon supplémentaire qui pourrait permettre de compenser pour un biais possible causé par l'échantillon longitudinal.

#### **4. ALLOCATION DE L'ÉCHANTILLON SUPPLÉMENTAIRE**

Les échéanciers étant serrés, il y avait peu de marge de manoeuvre pour déterminer comment allouer un échantillon supplémentaire. Différentes hypothèses ont été examinées; la sélection de tout un groupe de rotation de l'EPA, la sélection d'un sous-ensemble de RMR dans un ou deux groupes de rotation de l'EPA et différentes combinaisons d'un groupe complet avec un sous-ensemble de RMR. Les différents scénarios sont présentés dans le tableau 4.

**Tableau 4.** Différents scénarios d'allocation de l'échantillon supplémentaire (en termes de ménages)

	0 groupe de rotation complet	1 groupe de rotation complet
Aucun RMR supplémentaire		8,725
1 groupe 25 RMR	4,150	12,875
1 groupe 16 RMR	3,158	11,883
1 groupe 9 RMR	2,258	10,983
1 groupe 3 RMR	1,112	9,837
2 groupes 25 RMR	8,299	17,024
2 groupes 16 RMR	6,317	15,042
2 groupe 9 RMR	4,515	13240
2 groupe 3 RMR	2,224	10,949

Le groupe de 16 RMR correspond aux RMR pour lesquels l'EFC fourni ses estimations au niveau RMR, les 9 RMR représentent les RMR pour lesquels l'EPA a incorporé une stratification selon le revenu et les 3 RMR correspondent aux trois plus grandes villes. L'annexe 2 identifie les RMR appartenant à chacun des groupes.

L'étude est basée sur les hypothèses que l'on voulait assurer la qualité des estimations canadiennes de revenu individuel, familial et au niveau du nombre de personnes sous le seuil de pauvreté. Même si on désire des estimations nationales, il y aura probablement un besoin minimum d'estimations provinciales. Le fait de concentrer l'échantillon dans les RMR permettrait d'améliorer les estimations nationales avec un échantillon plus grand dans les gros RMR. De plus, l'expérience américaine suggère que l'érosion est plus importante dans les grandes villes. Par contre lorsqu'on analyse les taux de réponse longitudinal sur 3 ans de l'EDTR, présentés en annexe 3, les RMR n'ont pas un taux de réponse plus faible

que les CA. Également lorsque l'on compare les CV du tableau 2, il semble nécessaire de couvrir toutes les provinces si on veut assurer de bonnes estimations provinciales.

Le scénario privilégié serait le scénario où l'échantillon supplémentaire permettrait de choisir un groupe de rotation complet et un deuxième groupe de rotation qui ne couvrirait que les 9 RMR's. Ce scénario donne cependant une taille d'échantillon supérieure à ce qui était initialement prévu (soit 10,983 ménages au lieu de 9,400). Pour respecter les contraintes en terme de taille d'échantillon, il faut choisir entre un groupe de rotation complet (8,725) ou deux groupes de rotation dans les 25 RMR seulement (8,299). Comme les taux de réponse longitudinal sur 3 ans de l'EDTR ne sont pas plus faibles dans les RMR, nous recommandons d'utiliser un groupe de rotation complet pour la première année. En poursuivant les études sur l'érosion au cours des prochaines années, nous évaluerons s'il est préférable de concentrer l'échantillon dans les RMR.

## **5. DATE DE SÉLECTION DE L'ÉCHANTILLON ET UNITÉ D'ÉCHANTILLONNAGE LOGEMENT VS MÉNAGE**

Des comparaisons ont été faites pour tenter de voir quel serait le meilleur moment pour sélectionner l'échantillon supplémentaire. Différents scénarios ont été envisagés. D'abord l'échantillon supplémentaire peut être un échantillon de ménages choisi pendant l'année précédant la collecte, au plus tôt en janvier 1997, comme lors de l'introduction d'un nouveau panel de l'EDTR, au plus tard en août 1997, pour permettre la finalisation de l'application de collecte pour janvier 1998. L'échantillon supplémentaire pourrait également être un échantillon de logements choisi au moment de la collecte comme le fait actuellement l'EFC.

Le tableau de l'annexe 4 décrit en détail les avantages et désavantages de ces diverses options. Le principal désavantage de l'échantillon de ménages choisi en 1997 est qu'il nécessite du dépistage comme l'échantillon original. Le nombre d'individus à dépister est beaucoup plus faible pour l'échantillon choisi en août que pour celui choisi en janvier. Par contre l'échantillon choisi en janvier permet de faire de la rétro-information de données du travail de sorte à diminuer l'erreur de rappel et améliorer la qualité des données sur le travail. L'échantillon de logements a comme principal avantage d'assurer une meilleure représentativité transversale. Cette caractéristique en fait la meilleure approche d'un point de vue méthodologique puisque l'échantillon supplémentaire est sélectionné principalement dans ce but. Toutefois une étude des coûts reliés à cette approche (voir annexe 5) indique que l'échantillon devrait être réduit environ du tiers pour compenser les coûts supplémentaires reliés aux entrevues personnelles et au besoin de modifier l'application de collecte actuelle de l'EDTR. Il est donc proposé de choisir un échantillon de ménages pour la première année d'intégration avec la possibilité de modifier cette approche après avoir étudié de plus près la représentativité des panels de l'EDTR et considérant que l'application de collecte de l'EDTR devra de toute façon être modifiée dans les prochaines années.

## **6. CHOIX DE L'ÉCHANTILLON (GROUPE DE ROTATION)**

Considérant que l'échantillon supplémentaire sera un échantillon de ménages composé d'un groupe de rotation et choisi entre janvier et août 1997, il reste à déterminer quel groupe de rotation de l'EPA correspondra à l'échantillon supplémentaire de l'EDTR. En plus de considérer le problème de dépistage déjà mentionné dans la section précédente, le fardeau de réponse, le chevauchement entre les entrevues de l'EPA et celles de l'EDTR, le chevauchement par rapport au contenu et l'effet de grappe entre certains groupes de rotation sont des facteurs

déterminant dans le choix de l'échantillon. Le tableau de l'annexe 6 indique les différents échantillons possibles, les nombres 1 à 6 représentant les 6 groupes de rotation de l'EPA et les changements de couleur indiquant un renouvellement de l'échantillon pour ce groupe.

Les groupes 2 et 3 ont été écartés des choix possibles pour l'échantillon supplémentaire à cause de l'effet de grappe. Ce problème est engendré par l'approche utilisée par l'EPA pour remplacer l'échantillon. Lorsqu'un groupe de rotation a terminé ses 6 mois dans l'EPA, il est remplacé par un autre échantillon provenant généralement de la même grappe que le groupe sortant. Parmi tous les groupes de rotation remplacés dans une année, on estime qu'environ 90% sont remplacés par des échantillons de la même grappe. Pour l'EPA le problème de l'effet de grappe ne se pose pas puisqu'il y a seulement un groupe d'une même grappe dans l'échantillon à la fois. Pour l'échantillon transversal de l'EDTR qui sera composé de deux panels et de l'échantillon supplémentaire, il serait préférable que les échantillons ne proviennent pas des mêmes grappes. Comme le premier panel de l'EDTR provient de l'ancien design de l'EPA, les chances que les mêmes grappes soient considérées sont plutôt faibles. Toutefois le deuxième panel est composé du groupe de rotation se trouvant dans l'échantillon de l'EPA de août 1995 à janvier 1996 (groupe 2) et de celui dans l'échantillon de septembre 1995 à février 1996 (groupe 3). Il est donc préférable d'éliminer tous les échantillons provenant de groupes de rotation 2 et 3 pour éviter que l'échantillon supplémentaire ne soit concentré dans les mêmes grappes que le deuxième panel.

Certains groupes ont également été écartés parce que le fardeau de réponse était trop important et risquait d'avoir un effet sur le taux de réponse. Il s'agit des groupes de rotation déjà sollicités par l'EFC pour les revenus de 1997 et de certains groupes participant à plusieurs enquêtes supplémentaires de l'EPA (voir

annexe 6). Considérant que l'échantillon supplémentaire sera un échantillon de ménages qui doit par conséquent être sélectionné au plus tard en août 1997, le choix retenu est le groupe de rotation 4 qui sera dans l'EPA de avril 1997 à septembre 1997.

## **7. BESOINS À MOYEN TERME**

À moyen terme, une évaluation en profondeur de l'érosion de l'EDTR sera faite à mesure que les panels subséquents seront traités. A priori, il est peu probable que l'on ne voudra jamais laisser tomber l'option d'un échantillon supplémentaire pour maintenir la qualité des observations transversales, si on cherche à utiliser ces données comme sources de statistiques transversales officielles. De façon transversale, le panel va toujours souffrir d'un problème de représentativité. Il y a d'abord les ménages "d'immigrants" qui sont sous représentés ; mais on peut douter de la capacité d'un échantillon supplémentaire à bien cibler un si petit groupe. Plus important est cependant le problème d'érosion. On remarque que l'érosion est souvent plus importante dans certains groupes. Un groupe non étudié ici est les jeunes. Ils sont sous représentés lors des appariements aux données d'impôt et c'est un groupe où on observe une plus grande érosion. Il se peut éventuellement que la façon optimale d'assurer la représentativité transversale soit de réduire la taille des panels (par exemple en sous-échantillonnant les travailleurs payés dans certains groupes d'âge) afin d'augmenter la taille de l'échantillon supplémentaire pour ainsi aller chercher certains segments de la population qui sont sous représentés. Le temps et le manque de données longitudinales ne permettent pas de faire des inférences appropriées sur le problème pour le moment. Il ne permet surtout pas de quantifier l'impact d'une stratégie de sur-échantillonnage dans certains groupes donnés. Les recommandations sont donc de choisir un échantillon supplémentaire de façon simple à court terme et de refaire

des évaluations après la première année d'intégration. Lorsque les données du premier échantillon supplémentaire seront disponibles, de meilleures études comparatives entre la représentativité des panels vs un échantillon "à jour" pourront être effectuées et on pourra quantifier les impacts.

**ANNEXE 1 : CARACTÉRISTIQUES DES PERSONNES SELON LA  
RÉPONSE AUX INTERVIEWS : PANEL COMPLET DE  
L'ENQUÊTE SIPP DE 1984**

	Échantillon restreint (1)	Réponse à toutes les interviews (2)	Érosion (3)	Une interview manquée (4)	Au moins deux interviews manquées (5)
Nombre total	32 985	22 814	6 957	2 330	884
<b>Bureau régional :</b>					
Boston	7,0	7,2	7,4	5,1	3,7
New York	7,1	5,8	9,4	10,5	12,9
Philadelphie	10,3	11,0	7,6	9,8	13,6
Détroit	8,5	8,6	8,3	7,1	8,9
Chicago	7,6	8,7	5,4	5,3	4,1
Kansas City	8,4	9,5	6,4	5,1	3,4
Seattle	8,6	9,1	6,9	8,6	8,9
Charlotte	9,0	9,4	8,7	8,0	4,9
Atlanta	11,4	10,6	12,5	15,5	11,8
Dallas	9,5	8,5	11,4	12,2	12,2
Denver	5,8	5,7	6,3	5,2	4,6
Los Angeles	6,9	5,9	9,7	7,6	11,0

	Échantillon restreint (1)	Réponse à toutes les interviews (2)	Érosion (3)	Une interview manquée (4)	Au moins deux interviews manquées (5)
<b>Caractéristiques</b>					
<b>résidentielles :</b>					
Pas une SMSA *	25,2	27,4	20,1	22,4	17,2
SMSA : Moins de 100 000	1,2	1,2	1,3	1,0	0,6
SMSA : de 100 à 249 milliers	9,1	9,6	8,1	8,4	6,7
SMSA : de 250 à 499 milliers	9,3	9,7	8,9	7,6	8,5
SMSA : de 500 à 999 milliers	13,4	13,2	14,6	13,6	11,1
SMSA : de 1 à 2,9 millions	24,2	23,2	26,5	25,7	29,4
SMSA : de 3 à 14,9 millions	17,4	15,7	20,5	21,4	26,6
<b>Logement :</b>					
Maison ou app.	94,2	94,1	94,4	94,2	94,2
Tous les autres types de logement	5,8	5,9	5,6	5,8	5,8
<b>Logement :</b>					
Possédé/en cours d'achat	69,8	72,7	62,9	65,6	59,8
Loué au comptant	27,8	24,8	35,1	32,1	37,3
Occupé gratuitement	2,3	2,4	2,0	2,3	2,8

	Échantillon restreint (1)	Réponse à toutes les interviews (2)	Érosion (3)	Une interview manquée (4)	Au moins deux interviews manquées (5)
<b>Race :</b>					
Blanche	86,8	88,5	83,8	82,9	78,2
Noire	10,5	9,2	12,9	13,9	17,6
Amérindienne/Inuit/ toute race autochtone	0,4	0,3	0,7	0,5	0,9
Asiatique/ océanienne	2,2	2,0	2,6	2,7	3,3

\* SMSA : Standard Metropolitan Statistical Area (Zone statistique métropolitaine normalisée).

**ANNEXE 2 : DÉFINITION DES RMR**

	3 RMR	9 RMR	16 RMR	25 RMR
St John's			X	X
Halifax			X	X
St-John			X	X
Québec			X	X
Trois-Rivières				X
Sherbrooke				X
Montréal	X	X	X	X
Hull-Ottawa		X	X	X
Chicoutimi				X
Oshawa				X
Toronto	X	X	X	X
Hamilton		X	X	X
St-Cath/Niagara			X	X
London		X	X	X
Windsor				X
Kitchener			X	X
Sudbury				X
Thunder Bay				X
Winnipeg		X	X	X
Régina				X
Saskatoon				X
Calgary		X	X	X
Edmonton		X	X	X
Vancouver	X	X	X	
Victoria			X	X

### ANNEXE 3 : TAUX DE RÉPONSE LONGITUDINAL SUR TROIS ANS DE L'EDTR

RMR/AR	Nombre de répondants	Taux de réponse
<i>Manquant</i>	494	0.83
ST. JOHNS	241	0.84
010	46	0.88
011	8	0.53
015	93	0.86
025	20	0.87
105	244	0.89
110	64	0.72
HALIFAX	282	0.80
210	69	0.85
215	147	0.77
220	88	0.78
225	140	0.77
305	267	0.80
SAINT JOHN	196	0.78
320	208	0.87
328	37	0.65
330	20	0.95
335	67	0.92
403	20	0.91
404	66	0.77
406	65	0.73
CHICOUTIMI-	282	0.87
JONQUIÈRE		
410	9	0.82
411	20	0.77
412	48	0.76
QUÉBEC	228	0.75

RMR/AR	Nombre de répondants	Taux de réponse
SHERBROOKE	391	0.80
435	28	0.82
440	24	0.96
TROIS-RIVIÈRE	163	0.87
444	78	0.88
446	32	0.91
447	39	0.57
450	82	0.68
452	35	0.76
454	35	0.76
456	14	0.74
459	94	0.85
MONTRÉAL	625	0.77
465	45	0.94
468	11	0.85
475	61	0.77
480	48	0.77
485	81	0.80
501	106	0.74
502	2	0.67
HULL-OTTAWA	526	0.72
515	25	0.81
521	91	0.73
522	109	0.83
529	107	0.66
530	7	0.78
OSHAWA	269	0.79
TORONTO	699	0.76
HAMILTON	301	0.78
ST.CATHARINES/	426	0.80
NIAGARA		
KITCHENER	417	0.83
543	114	0.83

RMR/AR	Nombre de répondants	Taux de réponse
544	55	0.86
546	8	1.00
547	12	0.75
550	189	0.87
553	62	0.61
LONDON	322	0.78
556	13	0.76
557	51	0.94
WINDSOR	273	0.78
562	42	0.71
566	17	0.74
567	14	0.82
568	50	0.79
569	39	0.81
571	28	0.90
575	122	0.74
SUDBURY	252	0.84
582	52	0.84
584	14	0.93
585	7	0.70
586	17	0.71
590	58	0.81
THUNDER BAY	265	0.83
598	7	1.00
WINNIPEG	468	0.76
604	6	0.35
607	53	0.82
610	61	0.77
640	76	0.97
REGINA	336	0.86
710	41	0.87
715	74	0.76
720	63	0.86

RMR/AR	Nombre de répondants	Taux de réponse
SASKATOON	309	0.82
730	22	0.73
735	30	0.86
745	97	0.71
750	34	0.67
805	56	0.86
810	74	0.83
CALGARY	473	0.84
830	72	0.81
833	14	0.78
EDMONTON	550	0.79
840	36	0.77
845	1	1.00
850	44	0.83
860	37	0.80
865	19	0.86
905	42	0.71
913	53	0.83
915	64	0.76
918	25	0.81
925	86	0.77
930	124	0.79
932	137	0.76
VANCOUVER	407	0.74
VICTORIA	202	0.81
937	2	1.00
938	71	0.68
940	22	0.63
943	29	0.93
944	18	0.78
945	15	0.83
950	26	0.72
952	6	1.00

---

RMR/AR	Nombre de répondants	Taux de réponse
955	21	0.72
960	33	0.94
965	26	0.72
970	75	0.89
975	24	0.83
977	18	0.78
<i>Ne sait pas</i>	232	0.56
HORS RMR/AR	9659	0.84
TOTAL	24986	

**ANNEXE 4 : SÉLECTION DE L'ÉCHANTILLON SUPPLÉMENTAIRE  
SUPPOSANT LE MÊME CONTENU**

Grille de comparaison

	<b>ÉCHANTILLON DE MÉNAGES CHOISI EN JANVIER 1997</b>	<b>ÉCHANTILLON DE MÉNAGES CHOISI EN AOÛT 1997</b>	<b>ÉCHANTILLON DE LOGEMENTS CHOISI AU MOMENT DE LA COLLECTE</b>
<b>REPRÉSENTATIVITÉ DE L'ÉCHANTILLON</b>	(-) Les individus qui déménagent risquent d'être sous-représentés si les mesures prises pour le dépistage ne sont pas efficaces.  (-) Les immigrants de l'année 1997 ne peuvent pas être dans l'échantillon donc l'échantillon supplémentaire ne permettra pas de compenser la sous-couverture pour les immigrants.	(+-) La sous-représentativité des individus qui déménagent est moins importante puisqu'environ 2/3 des déménagements de l'année ont déjà eu lieu en août selon les résultats de l'EDTR 1994.  (+-) Les immigrants entrés au pays avant août 1997 pourraient se trouver dans l'échantillon.	(+) Bonne représentativité des individus qui déménagent en détectant les nouveaux arrivants dans les logements sélectionnés.  (+) L'échantillon supplémentaire couvrira les immigrants de l'année 1997.
<b>DÉPISTAGE</b>	(-) Nécessite plus de dépistage, donc un coût de collecte plus élevé et un taux de réponse plus faible. (+) Réduit l'erreur de rappel pour janvier 1997 car l'information de l'EPA de janvier est disponible pour les deux groupes.	(+-) Nécessite moins de dépistage puisqu'une bonne proportion des déménagements ont déjà eu lieu.	(+) Aucun dépistage nécessaire puisqu'on prend le ménage dans le logement lors de la sélection.
<b>ERREUR DE RAPPEL</b>		(-) Pas de rappel possible puisque l'application de l'EDTR ne permet pas d'utiliser l'information de l'EPA d'avril 1997 qui est disponible pour les deux groupes.	(-) Pas de rappel possible

	<b>ÉCHANTILLON DE MÉNAGES CHOISI EN JANVIER 1997</b>	<b>ÉCHANTILLON DE MÉNAGES CHOISI EN AOÛT 1997</b>	<b>ÉCHANTILLON DE LOGEMENTS CHOISI AU MOMENT DE LA COLLECTE</b>
<b>FARDEAU DE RÉPONSE</b>	(+) Délai de 12 (ou 11) mois entre le contact de l'EDTR et le dernier contact de l'EPA	(-) Délai de 4 (ou 3) mois seulement entre le dernier contact de l'EPA et le contact de l'EDTR. Une lettre incitative serait reçue peu de temps après l'EPA.  (-) Un des deux groupes de rotation aura déjà participé à l'EFC 1997, si on prend les deux groupes de rotation sortant.	(-) Les répondants seront sollicités pour l'EPA et pour l'EDTR-supplémentaire en janvier.
<b>COMPOSITION DE LA FAMILLE</b>	(+) La composition de la famille est définie au début de la période de référence donc on peut identifier les changements dans la composition de la famille pendant l'année de référence.	(+/-) La composition de la famille est inconnue pour les premiers mois de l'année de référence donc certains changements dans la composition de la famille ne pourront être identifiés.	(-) Les changements dans la composition de la famille ne pourront être identifier.
<b>ÉLIGIBILITÉ DES MEMBRES</b>	(+) L'éligibilité pourra être définie de la même façon que pour les deux panels.	(-) Plus difficile de définir quels sont les membres éligibles.	(-) Plus difficile de définir quels sont les membres éligibles.
<b>AUTRES</b>	(+) Approche plus cohérente avec les autres panels de l'EDTR puisque les échantillons sont choisis en janvier.		(-) Il faut modifier la procédure de contact de l'EDTR.

## **ANNEXE 5 : OPTIONS POUR L'ÉCHANTILLON SUPPLÉMENTAIRE DE 1998**

La présente note résume les répercussions sur la collecte des données et les estimations des coûts de plusieurs options pour l'échantillon supplémentaire de 1998. Fondamentalement, on compare un échantillon de logements axé sur les RMR et un échantillon de 9 000 ménages auxquels sont appliquées les méthodes actuelles de collecte de l'EDTR. L'utilisation d'un échantillon de logements nécessitera des visites sur place et la réécriture du module des contacts et des données démographiques de l'EDTR d'ici à janvier 1998.

Les programmes de toutes les enquêtes-ménages doivent être réécrits afin de pouvoir les exploiter dans l'environnement Windows et de résoudre les problèmes que pose l'année 2000. Avril 1999 est la date visée pour la mise en oeuvre sur le terrain. Pour toutes les enquêtes, la gestion des cas devra s'appuyer sur des interfaces normalisées et le «questionnaire» devra être exploitable dans l'environnement Windows. Les normes seront élaborées durant l'exercice 1997-1998 et ne seront pas établies au moment où nous serions censés effectuer la réécriture pour la collecte de janvier 1998 (cette réécriture devant débuter en avril 1997 au plus tard). Donc, nous serions obligés de remanier de nouveau le système pour la collecte de janvier 2000.

Par conséquent, si nous décidons d'utiliser un échantillon de logements, nous devons assumer les coûts de programmation et d'essai pour une application de collecte qui ne sera utilisée que deux ans.

La figure 1 résume les hypothèses et les scénarios opérationnels pour les deux options. Le tableau 1 donne les estimations de coûts et de taille d'échantillon pour

trois échantillons de logements englobant, respectivement, 9, 16 et 25 RMR, et pour deux échantillons de ménages, l'un avec un renouvellement complet et l'autre avec deux renouvellements axés sur les RMR ou les AR. Enfin, la figure 2 fournit une comparaison des avantages et des inconvénients des deux méthodes.

**Figure 1** : Hypothèses concernant la collecte

Échantillon de logements	Échantillon de ménages
Les logements vacants durant le mois de sélection nécessiteront une visite sur place.	Pas dans l'échantillon/sans objet.
Les logements vacants en janvier 1998 nécessiteront une visite sur place.	Probablement les cas «impossibles à dépister» de l'EDTR.
Les personnes ayant déménagé en janvier 1998 ne seront pas dépistées; une visite sur place sera nécessaire.	Les personnes ayant déménagé feront l'objet d'un dépistage par téléphone.
Les personnes ayant déménagé de janvier à mai devront être dépistées, ce qui implique une méthode de collecte différente pour chaque phase.	Les personnes ayant déménagé de janvier à mai devront être dépistées, donc la méthode de collecte sera la même pour chaque phase.
Réécriture importante du module de contact pour l'interview de janvier 1998.	Non nécessaire.
Étant donné la complexité de l'application de l'EDTR, on devra probablement créer des applications distinctes pour les cas supplémentaires et les cas longitudinaux en janvier.	Non nécessaire.
Réécriture importante du module de contact pour l'interview de janvier 2000 (toutes les enquêtes-ménages passent à l'environnement Windows).	Réécriture importante du module de contact pour l'interview de mai 1999 (toutes les enquêtes-ménages passent à l'environnement Windows).

**Tableau 1 Taille de l'échantillon et coûts**

	Échantillon de logements			Échantillon de ménages	
	9 RMR	16 RMR	25 RMR	1 groupe de rotation	sélection ciblée
Taille corrigée de l'échantillon <sup>1</sup>	4 924	6 889	9 050	9 000	9 000
Interviews sur place <sup>2</sup>	1 192	1 653	2 172	0	0
Coût de la collecte de janvier	102 676 \$	139 664 \$	180 347 \$	126 368 \$	126 368 \$
Coût de la collecte de mai	16 063	22 475	29 526	30 537	30 537
Programmation (une année uniquement) <sup>3</sup>	19 000	19 000	19 000	s.o.	s.o.
Question concernant les personnes-ressources, selon la remarque de l'EPA <sup>4</sup>	s.o.	s.o.	s.o.	1 500	3 000
<b>Coût total</b>	<b>137 739 \$</b>	<b>181 139 \$</b>	<b>228 873 \$</b>	<b>158 405 \$</b>	<b>159 905 \$</b>

1. Dénombrements des logements et des ménages fournis par la Division de la méthodologie, corrigés pour les logements vacants durant le mois de sélection.

2. Les estimations sont prudentes et fondées sur 8 % de logements vacants en 1997 et en 1998 (Méthodologie) et sur 8 % de déménagements de ménages entiers (vague 1 de l'EDTR). Le coût est de l'ordre de 30 \$ par cas comparativement à 14 \$ par interview téléphonique.

3. Hypothèse: cinq mois de travail d'un employé CS2.

4. Recommande de demander à deux groupes de rotation d'indiquer le nom d'une personne-ressource durant l'EPA de 1997. La Division de la méthodologie pourrait recommander de choisir soit l'échantillon de ménages avec renouvellement ciblé, soit l'échantillon de ménages avec un renouvellement complet, avec plus de renseignements sur l'érosion. D'ici à mars, nous devrions connaître les taux bruts de réponse pour le deuxième panel de la première vague (la grande inconnue). L'effet des travaux de préparation de la collecte devrait être à peu près le même pour les deux options.

**Figure 2** Avantages et inconvénients des deux options

	<b>Échantillon de logements</b>	<b>Échantillon de ménages</b>
<b>Avantages</b>	Le meilleur plan de sondage quant à la diminution du biais d'érosion.	Moins coûteux pour un échantillon de taille équivalente.
	Taux éventuellement plus élevé d'autorisation de l'utilisation des dossiers fiscaux.	Méthodes plus simples - maintien du plan fondamental de collecte de l'EDTR.
		Pas de gaspillage d'énergie dû à la réécriture du module de contact pour une utilisation limitée à deux ans.
		Modifications minimales de la formation et du manuel de l'intervieweur.
<b>Inconvénients</b>		
	Complique considérablement un système déjà complexe.	
	Envoie des messages contradictoires aux intervieweurs — le dépistage est important dans certains cas mais non dans d'autres.	

	Nécessite la modification du manuel de l'intervieweur; en outre, exigera une formation plus longue, sinon posera le risque d'une moins bonne connaissance des procédures longitudinales de l'EDTR.	
	Nécessite un compromis entre le coût et la taille de l'échantillon — effet sur les CV.	Ne compense pas aussi efficacement le biais d'érosion.
	Étant obligé de préparer des spécifications pour la réécriture du module de contact en 1997-1998, le personnel de l'EDTR disposera de moins de temps pour analyser les problèmes et optimiser les changements en prévision de la réécriture qui devra obligatoirement avoir lieu en 1998-1999.	
	Oblige à faire plus d'ajouts et de modifications au système de traitement.	Nécessite certaines modifications du système de traitement, mais moins nombreuses.
Inconnue	Effet sur les taux de réponse	Effet sur les taux de réponse



- Enquête sur les voyages (nov. ou déc. 1997, selon le groupe 4 ou 5)
  - Une personne par ménage
- Enquête sur l'importance de la faune (fév. 1997)
  - petite entrevue de moins de 5 minutes
- Enquête longitudinale sur les enfants (fév. ou mars 1997)
  - pour les ménages avec des enfants de 0 à 23 mois (environ 6% des ménages)

**Option 2: Les groupes 4 et/ou 5 choisis en août 1997**

3 à 4 mois entre la fin de l'EPA et le début de l'EDTR  
6 mois de chevauchement de contenu sur le travail avec l'EPA  
sélectionné 5 mois avant le contact (dépistage des déménagements)

fardeau supplémentaire: encore inconnu