



Guide de l'utilisateur des microdonnées

Enquête sur les approches en matière de planification des études

1999



Statistique
Canada

Statistics
Canada

Canada

Table des matières

1.0	Introduction	5
2.0	Contexte	7
3.0	Objectifs	9
4.0	Concepts et définitions	11
4.1	Concepts et définitions de l'Enquête sur la population active.....	11
4.2	Concepts et définitions de l'Enquête sur les approches en matière de planification des études	12
5.0	Méthodologie de l'enquête	15
5.1	Population visée.....	15
5.2	Plan de sondage	15
5.2.1	Stratification primaire	15
5.2.2	Types de régions	15
5.2.3	Stratification secondaire.....	16
5.2.4	Délimitation et sélection des grappes	16
5.2.5	Sélection des logements.....	17
5.2.6	Sélection des personnes.....	17
5.3	Taille de l'échantillon.....	17
5.4	Renouvellement de l'échantillon	18
5.5	Modifications au plan de l'Enquête sur la population active pour l'Enquête sur les approches en matière de planification des études	18
5.6	Taille de l'échantillon selon la province pour l'Enquête sur les approches en matière de planification des études	19
6.0	Collecte des données	21
6.1	Réalisation des interviews dans le cadre de l'Enquête sur la population active	21
6.2	Supervision et contrôle de qualité.....	21
6.3	Non-réponse à l'Enquête sur la population active	21
6.4	Modifications apportées à la collecte des données de l'Enquête sur les approches en matière de planification des études	22
6.5	Non-réponse à l'Enquête sur les approches en matière de planification des études.....	22
7.0	Traitement des données	23
7.1	Saisie des données.....	23
7.2	Vérification	23
7.3	Codage des questions ouvertes	23
7.4	Imputation	23
7.5	Création de variables dérivées	24
7.6	Pondération.....	26
7.7	Suppression des renseignements confidentiels	27
8.0	Qualité des données	29
8.1	Taux de réponse	29
8.2	Erreurs relatives à l'enquête	29
8.2.1	Base de sondage	30
8.2.2	Collecte des données	30
8.2.3	Traitement des données	31
8.2.4	Non-réponse	31
8.2.5	Mesure de l'erreur d'échantillonnage.....	33

9.0	Lignes directrices pour la totalisation, l'analyse et la diffusion de données	35
9.1	Lignes directrices pour l'arrondissement d'estimations	35
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	36
9.3	Définitions de types d'estimations : catégoriques et quantitatives	36
9.3.1	Estimations catégoriques	36
9.3.2	Estimations quantitatives	36
9.3.3	Totalisation d'estimations catégoriques	37
9.3.4	Totalisation d'estimations quantitatives	37
9.4	Lignes directrices pour l'analyse statistique	38
9.5	Lignes directrices pour la diffusion de coefficients de variation	39
9.6	Seuils pour la diffusion des estimations pour l'Enquête sur les approches en matière de planification des études	41
10.0	Tables de variabilité d'échantillonnage approximative	43
10.1	Comment utiliser les tables de coefficients de variation pour des estimations catégoriques	44
10.1.1	Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques	45
10.2	Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance	48
10.2.1	Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance	49
10.3	Comment utiliser les tables de coefficients de variation pour effectuer un test t	49
10.3.1	Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t	50
10.4	Coefficients de variation pour des estimations quantitatives	50
10.5	Tables des coefficients de variation	51
11.0	Pondération	53
11.1	Procédures de pondération pour l'Enquête sur la population active	53
11.2	Procédures de pondération pour l'Enquête sur les approches en matière de planification des études	54
12.0	Questionnaires	57
12.1	Le questionnaire de l'Enquête sur la population active	57
12.2	Le questionnaire de l'Enquête sur les approches en matière de planification des études	57
13.0	Clichés d'enregistrements à valeurs univariées	59
13.1	Unité d'analyse	59
13.2	Disponibilité de l'information sur les parents	59
13.3	Convention de désignation des variables	60

1.0 Introduction

Statistique Canada a mené l'Enquête sur les approches en matière de planification des études (EAPE) en octobre 1999 avec l'appui et la collaboration de Développement des ressources humaines Canada. Ce manuel a été produit pour faciliter la manipulation du fichier de microdonnées portant sur les résultats de l'enquête.

Toutes les questions concernant l'ensemble de données ou son utilisation devraient être adressées à :

Statistique Canada

Services à la clientèle

Centre de la statistique de l'éducation

Téléphone : (613) 951-7608 ou appelez sans frais : 1 800 307-3382

Télécopieur : (613) 951-9040

Courriel : educationstats@statcan.ca

2.0 Contexte

Développement des ressources humaines Canada (DRHC) a demandé à Statistique Canada de procéder à une enquête pour examiner comment les Canadiens préparent leurs enfants aux études postsecondaires. L'objectif était d'examiner les aspects financiers et « culturels » de la préparation. Le volet financier comprend les frais que prévoient les parents pour les études postsecondaires des enfants, ainsi que la somme et le genre d'épargnes réalisées à cette fin. Le volet culturel comprend certains points, notamment, l'engagement des parents dans la scolarité de leurs enfants et la prestation de ressources scolaires.

L'épargne pour les études postsecondaires n'est pas une activité restreinte aux parents d'enfants d'âge scolaire. De nombreux autres particuliers, par exemple les grands-parents, épargnent aussi pour les études postsecondaires d'un enfant. DRHC a décidé d'ajouter à l'enquête les ménages sans enfant, afin d'obtenir des données sur l'épargne aux fins des études pour tous les ménages au Canada. On a posé uniquement des questions sur l'épargne à ces ménages. L'ajout de ces ménages permet de formuler des énoncés sur toutes les épargnes aux fins des études au Canada, et non seulement sur les épargnes des ménages qui ont des enfants¹.

L'Enquête sur les approches en matière de planification des études de 1999 est la première enquête de Statistique Canada visant à obtenir de l'information détaillée sur les dispositions que prennent les Canadiens pour les études postsecondaires de leurs enfants. Étant donné l'importance croissante des études postsecondaires au Canada et des augmentations récentes des coûts de ces études, l'enquête donnera un premier aperçu important sur les dispositions que prennent les Canadiens pour les études postsecondaires de leurs enfants.

¹ Les données sur les ménages qui ont des enfants ne sont pas disponibles au fichier de microdonnées à grande diffusion, mais il est possible de les obtenir par totalisations personnalisées auprès de la Sous-section des services à la clientèle au Centre de la statistique de l'éducation.

3.0 Objectifs

L'objectif de l'Enquête sur les approches en matière de planification des études est de tracer un profil des deux moyens que les Canadiens peuvent adopter pour se préparer aux études postsecondaires de leurs enfants :

- 1) Préparation non financière - comment les parents sont-ils engagés dans les études et l'apprentissage actuels de leurs enfants?
- 2) Préparation financière - quelle est la sensibilisation au coût des études postsecondaires, quels sont les plans appliqués pour payer les études des enfants et quelles sont les épargnes antérieures et actuelles?

4.0 Concepts et définitions

Ce chapitre donne un aperçu des concepts et des définitions d'intérêt pour les utilisateurs. Les concepts et les définitions utilisés dans le cadre de l'Enquête sur la population active (EPA) sont décrits à l'intérieur de la section 4.1, tandis que ceux et celles propres à l'Enquête sur les approches en matière de planification des études (EAPE) sont fournis dans la section 4.2. Les utilisateurs sont priés de se reporter au chapitre 12.0 de ce document où figure une copie des formulaires d'enquête réellement employés.

4.1 Concepts et définitions de l'Enquête sur la population active

Situation vis-à-vis de l'activité

Désigne la situation de l'enquêté sur le marché du travail : un membre de la population active âgé de 15 ans et plus (à l'exclusion des pensionnaires des établissements) est soit **occupé**, soit **en chômage**, soit **inactif**.

Emploi

Les personnes occupées sont celles qui, au cours de la semaine de référence :

- a) ont fait un travail² quelconque dans le cadre d'un emploi ou dans une entreprise; ou
- b) avaient un emploi, mais n'étaient pas au travail à cause d'une maladie ou d'une incapacité, pour obligations personnelles ou familiales, pour des vacances, par suite d'un conflit de travail ou du fait de tout autre facteur (cela n'inclut pas les personnes mises à pied, celles qui étaient inactives entre deux emplois occasionnels et celles qui avaient un emploi devant commencer à une date ultérieure).

Chômage

Les chômeurs sont les personnes qui, au cours de la semaine de référence :

- a) avaient été mises à pied temporairement, mais s'attendaient à être rappelées au travail et étaient disponibles pour travailler; ou
- b) étaient sans emploi, avaient activement cherché un emploi au cours des quatre dernières semaines et étaient disponibles pour travailler³; ou
- c) devaient commencer un nouvel emploi dans quatre semaines ou moins à compter de la semaine de référence et étaient disponibles pour travailler.

² Travail inclut un travail quelconque contre rémunération ou en vue d'un bénéfice, c'est-à-dire tout travail rémunéré accompli pour un employeur ou à son propre compte, cela comprend aussi le travail familial non rémunéré, c'est-à-dire un travail (non rémunéré) qui contribue directement à l'exploitation d'une ferme, d'une entreprise ou d'un cabinet de professionnel appartenant à un membre apparenté du même ménage et exploité par lui. De telles activités peuvent inclure la tenue de livres, la vente de produits, le service aux tables, etc. Des tâches comme le ménage ou l'entretien ménager ne sont pas considérées comme un travail familial non rémunéré.

³ Les personnes sont considérées comme étant disponibles pour travailler si elles :

- i) ont déclaré qu'elles auraient pu travailler durant la semaine de référence si on leur avait offert un emploi convenable, ou encore si la raison pour laquelle elles ne pouvaient occuper un emploi était de nature temporaire, notamment du fait d'une maladie ou d'une incapacité, d'obligations personnelles ou familiales, parce qu'elles avaient déjà un emploi devant bientôt débiter, ou qu'elles étaient en vacances (avant 1997, les personnes en vacances n'étaient pas considérées comme étant disponibles).
- ii) étaient des étudiants à temps plein à la recherche d'un travail à temps partiel qui respectaient également la condition i) ci-dessus. Les étudiants à temps plein qui sont actuellement aux études et qui cherchent un travail à temps plein ne sont pas considérés comme étant disponibles au cours de la semaine de référence.

Inactifs

Les personnes inactives sont celles qui, considérant les conditions qui existaient sur le marché du travail dans leur région, ne désiraient ou ne pouvaient offrir ou fournir leurs services au cours de la semaine de référence. Autrement dit, elles n'étaient ni occupées ni en chômage.

Branche d'activité et de la profession

L'Enquête sur la population active fournit de l'information sur la branche d'activité et la profession des personnes occupées et en chômage et de celles inactives qui ont détenu un emploi au cours des 12 derniers mois. Depuis 1997, ces statistiques reposent sur Le système de classification des industries de l'Amérique du Nord (SCIAN) et La classification type des professions (CTP de 1991). Avant 1997, La classification type des industries de 1980 ainsi que La classification type des professions de 1980 étaient utilisées.

Semaine de référence

Semaine civile complète (du dimanche au samedi) sur laquelle porte, chaque mois, l'Enquête sur la population active. C'est habituellement la semaine où tombe le 15^e jour du mois. Les interviews sont réalisées durant la semaine suivante, appelée la semaine d'enquête, et la situation vis-à-vis de l'activité déterminée est celle de la semaine de référence.

Emploi à temps plein

L'emploi à temps plein tient compte des personnes qui travaillent habituellement 30 heures ou plus par semaine à leur emploi principal ou à leur unique emploi.

Emploi à temps partiel

L'emploi à temps partiel tient compte des personnes qui travaillent habituellement moins de 30 heures par semaine à leur emploi principal ou à leur unique emploi.

4.2 Concepts et définitions de l'Enquête sur les approches en matière de planification des études

Coûts des études

Droits de scolarité et autres coûts liés aux études (livres, frais du programme coopératif, fournitures de laboratoire, etc.).

Prêts aux étudiants subventionnés par le gouvernement

Prêts subventionnés par le gouvernement fédéral et/ou les gouvernements provinciaux, qui servent à payer les études postsecondaires.

Subventions ou bourses

Somme d'argent qu'un organisme donne à une personne pour payer ses études. Le bénéficiaire n'a pas à rembourser les subventions et les bourses.

Compte en fiducie au nom de

Sommes déposées en fiducie dans un compte de banque, de caisse de crédit ou de société de fiducie à l'intention du bénéficiaire, même s'il n'y a pas d'accord officiel de fiducie.

Frais de subsistance

Toutes dépenses engagées par un étudiant autre que les frais directs d'études. Les principales dépenses de subsistance seront probablement le loyer - l'hébergement. D'autres dépenses peuvent comprendre le transport, les aliments, les divertissements, les vêtements, etc.

Prêts non étudiants d'une institution financière

Comprend les prêts personnels, les marges de crédit personnelles et les prêts hypothécaires de premier ou de deuxième rang.

Autres prêts

Comprend les prêts provenant des employeurs, les avances salariales et les cartes de crédit.

Études postsecondaires

Tout genre d'études formelles après l'école secondaire, y compris le collège et l'université, ainsi que l'apprentissage, les programmes de formation professionnelle et technique, les collèges d'enseignement général et professionnel (cégeps au Québec) et d'autres programmes.

Régime enregistré d'épargne-études (REEE)

Moyen fiscalement avantageux d'épargner pour financer les études postsecondaires d'un enfant. Lorsque ce dernier entreprend des études postsecondaires, les gains du régime servent à payer les droits de scolarité ou d'autres dépenses pertinentes à ses études.

REEE - régimes collectifs

REEE qui fonctionnent selon le principe de la mise en commun. Le régime répartit les gains entre tous les bénéficiaires.

REEE - régime individuel

Les REEE sont vendus par des représentants qui gèrent chacun de leurs contrats REEE séparément.

Régime enregistré d'épargne-retraite (REER)

Programme d'accumulation du capital conçu pour encourager l'épargne aux fins de la retraite. Les cotisations sont déductibles d'impôt jusqu'à un certain point.

5.0 Méthodologie de l'enquête

Comme l'Enquête sur les approches en matière de planification des études (EAPE) a été menée en octobre 1999 auprès d'un sous-échantillon des logements inclus dans l'échantillon de l'Enquête sur la population active (EPA), son plan de sondage est donc étroitement lié à celui de l'EPA. Le plan de l'EPA est décrit brièvement à l'intérieur des sections 5.1 à 5.4⁴. Les sections 5.5 et 5.6 décrivent comment l'EAPE s'est écartée du plan de base de l'EPA d'octobre 1999.

5.1 Population visée

L'EPA est une enquête mensuelle réalisée auprès des ménages. Son échantillon est représentatif de la population canadienne civile non institutionnalisée de 15 ans et plus des 10 provinces du Canada. Sont spécifiquement exclus du champ de l'enquête les résidents du Yukon, des Territoires du Nord-Ouest et du Nunavut, les personnes qui vivent dans les réserves indiennes, les membres à temps plein des Forces armées canadiennes et les pensionnaires d'établissements. Réunies, ces personnes exclues de l'enquête représentent environ 2 % de la population de 15 ans et plus.

5.2 Plan de sondage

L'EPA a fait l'objet d'un remaniement poussé, dont le point culminant a été l'introduction du nouveau plan à la fin de 1994. L'échantillon de l'EPA repose sur un échantillonnage probabiliste stratifié faisant appel à un plan à plusieurs degrés à tous les stades de ce dernier. Les principes du plan sont les mêmes pour chaque province. Un diagramme résumant les stades du plan figure dans le document EPA_AnnexeA.pdf.

5.2.1 Stratification primaire

Les provinces sont divisées en régions économiques (RÉ) et en régions économiques d'assurance-emploi (RÉAE). Les RÉ sont des régions géographiques d'une structure économique plus ou moins homogène formées à la suite d'ententes fédérales-provinciales. Elles sont relativement stables au fil du temps. Les RÉAE sont également des régions géographiques et en gros ont la même taille et sont aussi nombreuses que les RÉ, mais ne sont pas définies de la même façon. On produit des estimations de la population active pour les RÉAE aux fins de Développement des ressources humaines Canada.

Les intersections des régions des deux types susmentionnés forment le premier niveau de stratification pour l'EPA. On traite ces intersections de RÉ et de RÉAE comme des strates primaires et on effectue une stratification plus poussée à l'intérieur d'elles (voir la section 5.2.3). À noter que la stratification incluse dans le plan actuel de l'EPA respecte aussi un troisième ensemble de régions, les régions métropolitaines de recensement (RMR), puisque chaque RMR est également une RÉAE.

5.2.2 Types de régions

Les strates primaires (les intersections de RÉ et de RÉAE) sont en outre désagrégées en trois types de régions : rurales, urbaines et éloignées. Les régions rurales et urbaines sont généralement fondées sur les définitions du recensement de régions urbaines et de régions rurales, à quelques exceptions près destinées à permettre la formation de strates

⁴ Une description détaillée du plan de l'EPA figure dans la publication de Statistique Canada intitulée *Méthodologie de l'Enquête sur la population active du Canada*, N^o 71-526-XPB au catalogue.

dans certaines régions. Les régions urbaines vont des RMR les plus grandes jusqu'aux villages les plus petits classés suivant le Recensement de 1991 dans la catégorie des régions urbaines (de 1 000 habitants ou plus), tandis que les régions rurales se composent des régions non désignées régions urbaines ni régions éloignées.

Toutes les régions urbaines sont, en outre, subdivisées en deux types : celles pour lesquelles on utilise une liste d'appartements et une base aréolaire et celles pour lesquelles on emploie uniquement une telle base.

Environ 1 % de la population visée par l'EPA se trouve dans des régions éloignées des provinces qui sont moins accessibles que d'autres régions pour les intervieweurs affectés à l'EPA. À des fins administratives, cette portion de la population est échantillonnée séparément à l'aide de la liste des régions éloignées. Certaines populations, non rassemblées dans des endroits comptant 25 habitants ou plus, sont exclues de la base de sondage.

5.2.3 Stratification secondaire

Dans les régions urbaines où le nombre d'immeubles d'appartements est suffisamment grand, les strates sont subdivisées en listes d'appartements et en bases aréolaires. Une liste d'appartements est un registre tenu à jour dans les 18 plus grands centres du Canada. Cela vise à assurer une meilleure représentation des résidents d'appartements à l'intérieur de l'échantillon et à réduire l'effet de croissance dans les grappes, attribuable à la construction de nouveaux immeubles d'appartements. Dans les principaux centres, les strates d'appartements sont, en outre, subdivisées en strates de faibles revenus et en strates ordinaires.

Lorsque cela est possible et/ou nécessaire, la liste des régions urbaines est, en plus, subdivisée en strates ordinaires, en strates de revenus élevés et en strates de faible densité de population. La plupart des régions urbaines font partie des strates urbaines ordinaires, qui, en fait, englobent la majorité de la population canadienne. Les strates de revenus élevés se trouvent dans les principales régions urbaines, tandis que les strates urbaines de faible densité se composent de petites villes géographiquement dispersées.

Dans les régions rurales, la densité de population peut varier grandement, c'est-à-dire aller d'une densité de population relativement élevée à une faible densité de population, ce qui entraîne la formation de strates reflétant ces variations. Les différentes stratégies de stratification pour les régions rurales ont été fondées non seulement sur la concentration de la population, mais également sur le coût-efficacité et les contraintes auxquelles les intervieweurs sont confrontés.

Dans chaque province, le nombre de peuplements éloignés échantillonnés est proportionnel au nombre de logements, sans autre stratification. On sélectionne les logements à l'aide d'une méthode d'échantillonnage systématique dans chacun des endroits échantillonnés.

5.2.4 Délimitation et sélection des grappes

On ne sélectionne pas directement les ménages à l'intérieur des strates finales. On divise plutôt chaque strate en grappes, puis on sélectionne un échantillon de grappes à l'intérieur de la strate. On échantillonne ensuite les logements à partir des grappes sélectionnées. On utilise différentes méthodes pour définir les grappes, suivant le type de strate.

À l'intérieur de chaque strate urbaine incluse sur la liste des régions urbaines, on forme un certain nombre de groupes géographiquement continus de logements, ou grappes, à partir des chiffres du Recensement de 1991. Ces grappes sont généralement un ensemble d'un ou de plusieurs îlots, ou côtés d'îlot. La sélection d'un échantillon de grappes (toujours six ou un multiple de six grappes) à partir de chacune de ces strates secondaires représente le premier degré d'échantillonnage dans la plupart des régions urbaines. Dans certaines autres régions urbaines, on utilise comme grappes des secteurs de dénombrement (SD) du recensement. Dans les strates urbaines de faible densité de population, on suit un plan à trois degrés ou stades. Aux termes de ce plan, on échantillonne deux villes à l'intérieur d'une strate, puis 6 ou 24 grappes à l'intérieur de chaque ville.

Pour les strates d'appartements des régions urbaines, plutôt que de définir des grappes, on utilise l'immeuble d'appartements comme unité primaire d'échantillonnage. On échantillonne les immeubles d'appartements à partir de la liste, la probabilité d'échantillonnage étant proportionnelle au nombre d'unités que renferme chaque immeuble.

À l'intérieur de chacune des strates secondaires des régions rurales, on effectue, si nécessaire, une autre stratification afin de refléter les différences entre un certain nombre de caractéristiques socio-économiques de chaque strate. À l'intérieur de chaque strate d'une région rurale, on échantillonne comme grappes six SD ou deux ou trois groupes de SD.

5.2.5 Sélection des logements

Dans les trois types de régions que sont les régions urbaines, rurales et éloignées, des enquêteurs itinérants visitent premièrement les grappes sélectionnées, puis on dresse une liste de tous les logements privés faisant partie des grappes. On sélectionne ensuite à partir de cette liste un échantillon de logements. Le rendement de l'échantillon dépend du type de strate. Sur la liste de régions urbaines, par exemple, les rendements de l'échantillon sont de six ou de huit logements, selon la taille du centre. Le rendement de chaque grappe inscrite sur la liste d'appartements d'une région urbaine, est de cinq logements, tandis que dans les régions rurales et dans les parties de centres formées par des SD le rendement de chaque grappe est de dix logements. Dans toutes les grappes, on échantillonne systématiquement les logements, ce qui représente le dernier degré d'échantillonnage.

5.2.6 Sélection des personnes

On recueille des renseignements démographiques sur tous les membres du ménage pour lesquels le logement sélectionné constitue le lieu de résidence habituel. Les renseignements obtenus dans le cadre de l'EPA concernent tous les membres civils du ménage âgés de 15 ans ou plus. Pour les personnes âgées, c'est-à-dire celles de 70 ans et plus, on réduit le fardeau de la personne en réutilisant leurs réponses à l'interview initiale au cours des cinq mois suivants de l'enquête.

5.3 Taille de l'échantillon

On détermine la taille de l'échantillon des personnes admissibles dans le cadre de l'EPA de façon à respecter les exigences en matière de précision statistique établies pour diverses caractéristiques de la population active aux niveaux provincial et infraprovincial et à répondre aux besoins des administrations fédérale, provinciales et municipales et à ceux d'une foule d'autres utilisateurs de données.

L'échantillon mensuel de l'EPA se compose d'environ 60 000 logements. Après en avoir exclu les logements trouvés vacants, les logements démolis ou ceux convertis à des fins non résidentielles, ceux n'abritant que des personnes inadmissibles, les logements en construction et les logements saisonniers, il reste à peu près 54 000 logements occupés par une ou par plusieurs personnes admissibles. On recueille de l'information aux fins de l'EPA à partir de ces logements sur environ 102 000 civils âgés de 15 ans et plus.

5.4 Renouvellement de l'échantillon

L'EPA utilise un plan de sondage avec renouvellement de panel, suivant lequel les ménages sélectionnés restent dans l'échantillon pendant six mois consécutifs. L'échantillon complet est formé de six sous-échantillons ou panels représentatifs. Chaque mois, on remplace le panel qui fait partie de l'échantillon depuis six mois. Les ménages qui cessent de faire partie de l'échantillon sont remplacés par d'autres ménages du même secteur ou d'un secteur comparable. Il en résulte un chevauchement des cinq sixièmes de l'échantillon d'un mois à l'autre, ce qui procure un plan de sondage efficace pour estimer les variations d'un mois à l'autre. La rotation après six mois évite d'imposer un fardeau trop lourd aux enquêtés des ménages sélectionnés aux fins de l'enquête.

Grâce à la caractéristique que constituent les groupes de renouvellement, on peut mener facilement des enquêtes supplémentaires à l'aide du plan de l'EPA, mais en utilisant un échantillon moins complet.

5.5 Modifications au plan de l'Enquête sur la population active pour l'Enquête sur les approches en matière de planification des études

On a utilisé pour l'Enquête sur les approches en matière de planification des études cinq des six groupes de renouvellement inclus dans l'échantillon de l'EPA d'octobre 1999. Contrairement à l'EPA, pour laquelle des données sont recueillies pour tous les membres admissibles d'un ménage, l'EAPE n'a donné lieu à la cueillette d'information qu'auprès d'un seul membre du ménage sur les enfants de 18 ans ou moins dans le ménage, et sur les activités d'épargne du ménage.

Après avoir achevé l'entrevue de l'Enquête sur la population active, les intervieweurs ont demandé à parler au répondant de l'Enquête sur les approches en matière de planification des études. S'il n'y avait pas d'enfants de 18 ans ou moins dans le ménage, le répondant de l'EAPE était le même que celui de l'EPA. Dans les ménages qui avaient des enfants de 18 ans ou moins, le répondant à l'EAPE était celui qui avait davantage de connaissances des enfants du ménage et des plans établis pour les études postsecondaires des enfants.

5.6 Taille de l'échantillon selon la province pour l'Enquête sur les approches en matière de planification des études

Le tableau qui suit montre le nombre de ménages et d'enfants inclus dans les groupes de renouvellement échantillonnés de l'EPA qui étaient admissibles pour le supplément de l'Enquête sur les approches en matière de planification des études.

Provinces	Taille de l'échantillon des ménages *	Taille de l'échantillon des enfants **
Terre-Neuve-et-Labrador	1 465	965
Île-du-Prince-Édouard	1 069	670
Nouvelle-Écosse	2 650	1 557
Nouveau-Brunswick	2 300	1 466
Québec	8 261	4 819
Ontario	12 404	8 143
Manitoba	3 071	1 929
Saskatchewan	3 238	2 211
Alberta	3 261	2 392
Colombie-Britannique	3 880	2 314
Canada	41 599	26 466

* Seuls les ménages de l'EPA des groupes de renouvellement 1, 2, 3, 5 et 6 qui avaient autorisé la communication téléphonique étaient admissibles pour le supplément de l'EAPE.

** Les données ont été obtenues pour tous les enfants des ménages admissibles au supplément de l'EAPE, jusqu'à un maximum de trois enfants par ménage. Dans les ménages qui en comptaient davantage, les trois enfants de l'échantillon ont été sélectionnés au hasard.

6.0 Collecte des données

La collecte des données aux fins de l'Enquête sur la population active (EPA) a lieu tous les mois durant la semaine qui suit la semaine de référence de l'EPA. La semaine de référence correspond habituellement à la semaine où tombe le 15 du mois.

6.1 Réalisation des interviews dans le cadre de l'Enquête sur la population active

Les intervieweurs de Statistique Canada sont des employés embauchés et formés pour mener l'EPA et d'autres enquêtes auprès des ménages. Chaque mois, ils communiquent avec les logements de l'échantillon pour recueillir l'information requise au sujet de la population active. Chaque intervieweur entre en contact avec environ 75 logements par mois.

Dans le cas des logements nouvellement incorporés à l'échantillon, l'intervieweur mène habituellement une interview sur place assistée par ordinateur (IPAO). L'intervieweur recueille premièrement des renseignements socio-démographiques sur chaque membre du ménage, puis recueille des renseignements relatifs à la population active concernant tous les membres du ménage qui sont âgés de 15 ans et plus et ne sont pas membres des Forces armées. Pourvu qu'il y ait un appareil téléphonique dans le logement et que l'autorisation en ait été accordée, les interviews subséquentes sont réalisées par téléphone, par moyen d'interview téléphonique assistée par ordinateur (ITAO), provenant d'une unité centralisée là où les cas sont assignés au hasard aux intervieweurs. Ainsi, environ 85 % de l'ensemble des ménages sont interviewés par téléphone. Lors de ces interviews mensuelles subséquentes, l'intervieweur confirme les renseignements socio-démographiques obtenus le premier mois, puis il recueille les renseignements relatifs à la population active pour le mois courant.

Pour chaque logement, l'intervieweur obtient habituellement les renseignements relatifs à tous les membres du ménage auprès d'un membre bien informé du ménage. Appelée réponse « par procuration », cette façon de procéder est appliquée parce qu'il serait trop coûteux et fastidieux de faire plusieurs visites ou appels pour obtenir l'information directement auprès de chaque répondant. Environ 65 % des renseignements recueillis le sont de cette manière.

Lorsque, pendant la période de six mois où un logement fait habituellement partie de l'échantillon, un ménage entier déménage et est remplacé par un nouveau ménage, on recueille des renseignements sur le nouveau ménage pendant le reste de la période de six mois.

À la fin des interviews mensuelles menées pour l'EPA, les intervieweurs présentent le questionnaire d'enquête supplémentaire, le cas échéant, qu'il faut faire remplir à certains ou à la totalité des membres du ménage ce mois-là.

6.2 Supervision et contrôle de qualité

Tous les intervieweurs de l'EPA travaillent sous la supervision d'un groupe d'intervieweurs principaux. Ceux-ci ont pour responsabilité de s'assurer que les intervieweurs connaissent bien les concepts de l'EPA, les procédures qui y sont associées et ses nombreuses enquêtes supplémentaires. Ils ont aussi pour fonction de contrôler régulièrement le travail des intervieweurs et d'examiner les documents complétés par ces derniers. Les intervieweurs principaux sont, à leur tour, supervisés par les gestionnaires du programme de l'EPA, affectés dans chacun des bureaux régionaux de Statistique Canada.

6.3 Non-réponse à l'Enquête sur la population active

Les intervieweurs ont pour consigne de faire tous les efforts raisonnables pour obtenir des interviews aux fins de l'EPA avec des membres des ménages admissibles. Lorsqu'une personne

refuse au départ de participer à l'EPA, le bureau régional envoie à l'adresse du logement une lettre soulignant l'importance que revêtent l'enquête et la collaboration du ménage. Cette lettre est suivie d'un deuxième appel ou d'une deuxième visite de l'intervieweur. Dans les cas où la visite ou l'appel de l'intervieweur survient à un moment inopportun, on fixe un rendez-vous pour une visite de rappel à un moment plus opportun. Lorsqu'il n'y a personne dans le logement, l'intervieweur rappelle à de nombreuses reprises. Quelles que soient les circonstances, on ne remplace jamais des logements échantillonnés par d'autres logements en cas de non-réponse au questionnaire de l'EPA.

Chaque mois, après qu'on a déployé tous les efforts voulus pour réaliser les interviews, il reste un petit nombre de ménages qui n'ont pas répondu. Pour les ménages non-répondants et pour lesquels on a obtenu le mois précédent des renseignements aux fins de l'Enquête sur la population active, on reporte et on utilise ces renseignements à titre de données de l'enquête du mois courant. On ne collecte pas de données sur ces ménages dans le cadre d'une enquête supplémentaire.

6.4 Modifications apportées à la collecte des données de l'Enquête sur les approches en matière de planification des études

Étant donné les contraintes opérationnelles pendant la collecte des données, l'Enquête sur les approches en matière de planification des études (EAPE) s'est déroulée comme une enquête papier et crayon. Les étiquettes identifiant les particuliers sélectionnés ont été produites au bureau central avant la semaine de l'enquête et appliquées ensuite aux questionnaires dans les bureaux régionaux de Statistique Canada. Si le répondant à l'EAPE n'était pas le même que celui de l'EPA, l'intervieweur l'indiquait sur le questionnaire.

L'EAPE a ciblé un particulier par ménage. S'il n'y avait pas d'enfants de 18 ans ou moins dans le ménage, le répondant à l'EAPE était le même que celui de l'EPA. Dans les ménages qui avaient des enfants de 18 ans ou moins, le répondant à l'EAPE était celui qui connaissait le mieux les enfants du ménage et les plans appliqués pour leurs études postsecondaires.

À la fin de l'interview menée pour l'Enquête sur la population active, l'intervieweur demandait à parler au répondant de l'EAPE s'il n'était pas le même que celui de l'EPA. Si le répondant à l'EAPE n'était pas disponible, l'intervieweur fixait un moment opportun pour lui téléphoner. La période de collecte des données a été prolongée d'une semaine pour permettre aux intervieweurs de communiquer avec les répondants à l'EAPE.

L'EAPE a recueilli de l'information sur chaque enfant dans le ménage jusqu'à concurrence de trois enfants. Dans les ménages qui ont plus de trois enfants de 18 ans ou moins, les intervieweurs devaient sélectionner manuellement trois enfants à l'aide de la grille de l'étiquette du questionnaire.

6.5 Non-réponse à l'Enquête sur les approches en matière de planification des études

Pour les ménages ayant répondu au questionnaire de l'EPA, l'étape suivante de la collecte des données consistait à faire remplir le questionnaire de l'Enquête sur les approches en matière de planification des études. Au total, 41 599 ménages étaient admissibles à l'enquête supplémentaire. L'EAPE a été achevée dans 36 130 de ces ménages pour un taux de réponse de 86,9 %. Dans le cas des ménages admissibles, il y avait 26 466 enfants de 18 ans ou moins. Les intervieweurs ont obtenu de l'information sur 20 353 de ces enfants pour un taux de réponse de 76,9 %. De plus amples renseignements sur les taux de réponse sont présentés au chapitre 8.0 (Qualité des données).

7.0 Traitement des données

Le principal produit de l'Enquête sur les approches en matière de planification des études (EAPE) est un fichier de microdonnées « épuré ». Ce chapitre présente un bref résumé des phases de traitement inhérentes à la production de ce fichier.

7.1 Saisie des données

La saisie des données de l'enquête s'est effectuée à l'aide des installations de saisie de données situées dans chacun des bureaux régionaux de Statistique Canada. Au cours de ce processus, tout document contenant au moins un poste rempli par un intervieweur a été saisi et une version non vérifiée de l'enregistrement informatique a été transmise à Ottawa pour traitement plus poussé. On a saisi et transmis au total 42 425 documents pour l'enquête.

7.2 Vérification

La première étape du traitement d'enquête effectué au bureau central a été de remplacer les valeurs « hors limites » incluses dans le fichier de données par des blancs. Ce processus a été conçu pour faciliter les vérifications ultérieures.

Les erreurs dans le déroulement du questionnaire, où l'on a relevé des questions qui ne s'appliquaient pas au répondant (et auxquelles on n'aurait donc pas dû répondre) renfermaient des réponses constituaient le premier type d'erreurs traitées. Dans ces cas, une vérification par ordinateur a éliminé automatiquement les données superflues en suivant l'ordre du questionnaire dicté par les réponses à des questions antérieures et subséquentes, parfois.

Le second type d'erreurs traitées avait trait à un manque d'information dans les questions pour lesquelles le répondant aurait dû répondre. Pour ce type d'erreur, un code de non-réponse ou « non déclaré » était attribué au poste.

Des corrections ont aussi été apportées pour garantir que les données sur le nombre, l'âge et le sexe des enfants des ménages obtenues lors de l'EAPE correspondent à celles de l'Enquête sur la population active (EPA).

7.3 Codage des questions ouvertes

Quelques données élémentaires ont été consignées sur le questionnaire par les intervieweurs sous forme de questions ouvertes. On a inclus dans l'enquête au total sept questions partiellement ou complètement ouvertes. C'étaient des postes qui se rattachaient aux raisons pour lesquelles un enfant n'avait pas été à l'école en 1998-1999, à l'ethnicité et aux langues parlées dans le ménage.

7.4 Imputation

L'imputation est le processus qui fournit des valeurs valides concernant les variables qui ont été retenues pour être modifiées, soit en raison de renseignements invalides, soit en raison de renseignements manquants. Les nouvelles valeurs sont établies de façon à préserver la structure sous-jacente des données et à garantir que les enregistrements qui en résultent passeront tous les contrôles requis. En d'autres mots, l'objectif n'est pas de reproduire les véritables valeurs des microdonnées mais plutôt d'établir des enregistrements intrinsèquement cohérents qui permettront de produire de bonnes estimations agrégées.

Nous pouvons faire la distinction entre trois types de non-réponses. La non-réponse complète est lorsque le répondant ne fournit pas le nombre minimal de questions. Ces enregistrements sont supprimés et seront prises en compte lors du processus de pondération (voir le chapitre 11.0). La non-réponse est lorsque le répondant ne fournit pas une réponse à une question mais qu'il passe à la question suivante. Ces non-réponses sont habituellement traitées en utilisant le code « non déclaré » ou elles sont imputées. Enfin, la non-réponse partielle est lorsque le répondant fournit le nombre minimal de questions mais ne termine pas l'interview. Ces enregistrements pourraient être traités comme étant non-réponse complète ou étant non-réponse multiple.

Dans le cas de l'EAPE de 1999, l'imputation par enregistrement donneur a été utilisée pour fournir les éléments d'information manquants à six questions clés, en raison d'une non-réponse à une question. De plus amples renseignements sur le processus de l'imputation sont donnés au chapitre 8.0 (Qualité des données).

7.5 Création de variables dérivées

Un certain nombre de données élémentaires incluses dans le fichier de microdonnées ont été calculées en combinant des postes sur le questionnaire pour faciliter l'analyse des données. Voici quelques exemples de variables dérivées qui ont été créées.

Variables dérivées de l'enfant

Le fichier de microdonnées à grande diffusion de l'enfant et le fichier maître de l'enfant :

Statut d'épargne du ménage (HSAVST)

- Épargnant, enfants entre 0 et 18 ans dans le ménage
- Aucune épargne, enfants entre 0 et 18 ans dans le ménage

Statut des enfants à qui des économies sont destinées (HSAVST2)

- Épargne pour des enfants vivant à l'intérieur du ménage seulement
- Épargne pour des enfants vivant à l'extérieur du ménage seulement
- Épargne pour des enfants vivant à l'intérieur et à l'extérieur du ménage
- Aucune épargne

Statut de Régime enregistré d'épargne-études (REEE) du ménage (HRESPST)

- Ménage épargnant et utilisant un REEE
- Ménage épargnant mais n'utilisant pas un REEE
- Aucune épargne

Nombre total d'enfants vivant à l'intérieur et à l'extérieur du ménage à qui des épargnes sont destinées (TCHSAVAL)

Nombre total d'enfants vivant à l'intérieur du ménage à qui des épargnes sont destinées (TCHSAVIN)

Le fichier de microdonnées à grande diffusion de l'enfant seulement :

Langue principale parlée dans le ménage (MAINLANG)

- Anglais uniquement
- Français uniquement
- Autre uniquement
- Plus qu'une langue

- Autre(s) langue(s) parlées dans le ménage (OTHLANG)
- Aucune autre langue n'est parlée dans ce ménage
 - D'autres langues sont parlées dans ce ménage

Revenu du ménage (INCOMEDV)

- 0 \$ à 9 999 \$
- 10 000 \$ à 14 999 \$
- 15 000 \$ à 19 999 \$
- 20 000 \$ à 29 999 \$
- 30 000 \$ à 39 999 \$
- 40 000 \$ à 49 999 \$
- 50 000 \$ à 59 999 \$
- 60 000 \$ à 79 999 \$
- 80 000 \$ ou plus

Le fichier maître de l'enfant seulement :

Total des économies du ménage en 1998 (HSAV1998)

Total des économies du ménage en 1998 pour les enfants vivant à l'intérieur du ménage seulement (HSAV98IN)

Total des économies du ménage en 1999 (HSAV1999)

Total des économies du ménage en 1999 pour les enfants vivant à l'intérieur du ménage seulement (HSAV99IN)

Total des économies du ménage prévues pour le reste de 1999 (HSAVRT99)

Total des économies du ménage prévues pour le reste de 1999 pour les enfants vivant à l'intérieur du ménage seulement (HSAVR99I)

Total des économies du ménage (HSAVTOT)

Total des économies du ménage pour les enfants vivant à l'intérieur du ménage seulement (HSAVTOTI)

Total de l'épargne en utilisant un REEE (HRESPSAV)

Total de l'épargne en utilisant un REEE pour les enfants vivant à l'intérieur du ménage seulement (HRESPSIN)

Langue principale parlée dans le ménage (MAINLANG)

- Anglais uniquement
- Français uniquement
- Autre uniquement
- Anglais et français uniquement
- Anglais et autre
- Français et autre
- Anglais, français et autre

Revenu du ménage (INCOMEDV)

- Aucun revenu
- 1 \$ à 4 999 \$
- 5 000 \$ à 9 999 \$
- 10 000 \$ à 14 999 \$
- 15 000 \$ à 19 999 \$
- 20 000 \$ à 29 999 \$
- 30 000 \$ à 39 999 \$
- 40 000 \$ à 49 999 \$
- 50 000 \$ à 59 999 \$
- 60 000 \$ à 79 999 \$
- 80 000 \$ ou plus

Il existe également un nombre de variables dérivées qui ont été créées à la demande de Développement et ressources humaines Canada (DRHC) et qui sont disponibles sur le fichier maître et le fichier partagé des microdonnées. Le premier type de variable dérivée créée, détermine si un montant d'argent équivalent a été épargné pour tous les enfants du ménage. Par exemple, dans un ménage comptant trois enfants, le répondant a déclaré qu'en 1998 les membres du ménage avaient épargné 500 \$ pour le premier enfant, 500 \$ pour le second et 500 \$ pour le troisième. Ce type de variable dérivée a été créée pour :

- Montants épargnés en 1998 (DV32)
- Montants épargnés en 1999 (DV33)
- Total des épargnes (DV35)
- Épargnes dans un REEE (DV40)

Le deuxième type de variable a été créée pour repérer les dossiers d'enfants ayant des montants d'épargne non concordants. DRHC a demandé que ces irrégularités soient conservées sur le fichier maître et le fichier partagé des microdonnées. Ce type de variable dérivée a été créée pour vérifier que :

- Le montant d'argent total épargné pour un enfant soit inférieur ou égal à la somme des montants épargnée en 1998 et en 1999.
- Le montant d'argent épargné prévu d'ici à ce que l'enfant commence ses études postsecondaires soit inférieur ou égal au montant d'argent total déjà épargné.
- Le montant d'argent total épargné pour un enfant soit inférieur ou égal au montant épargné dans un REEE.

7.6 Pondération

Le principe qui sous-tend une estimation pour un échantillon probabiliste comme celui de l'EPA veut que chacune des personnes incluses dans l'échantillon « représente », en plus d'elle-même, plusieurs autres personnes qui en sont exclues. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne incluse dans l'échantillon représente 50 membres de la population.

La phase de la pondération est une étape où l'on calcule ce nombre (ou poids) pour chaque enregistrement. Ce poids, qui figure dans le fichier de microdonnées, **doit** servir à calculer des estimations significatives à partir de l'enquête. Si, par exemple, le nombre d'enfants qui ont des

épargnes pour leurs études postsecondaires, doit être estimé, cette opération s'effectue en sélectionnant les enregistrements se référant aux personnes incluses à l'intérieur de l'échantillon qui présentent cette caractéristique et en additionnant les poids inscrits dans ces enregistrements.

Le chapitre 11.0 renferme des détails au sujet de la méthode utilisée pour calculer ces poids.

7.7 Suppression des renseignements confidentiels

Il convient de souligner que les fichiers de microdonnées « à grande diffusion » décrits ci-dessus diffèrent sous un certain nombre d'aspects importants des fichiers « maîtres » de l'enquête que conserve Statistique Canada. Ces différences sont le résultat de mesures prises pour protéger l'anonymat des répondants à un questionnaire d'enquête. Les utilisateurs ayant besoin d'avoir accès à de l'information exclue des fichiers de microdonnées peuvent acheter des totalisations spéciales. Les estimations produites seront communiquées à l'utilisateur, sous réserve du respect des lignes directrices pour l'analyse et la diffusion dont le chapitre 9.0 de ce document fournit un aperçu.

Identificateurs géographiques

Le fichier maître de données de l'enquête inclut des identificateurs géographiques explicites pour la région, la province, la région économique et la région métropolitaine de recensement. Il est aussi possible d'obtenir, lorsque les tailles de l'échantillon le permettent, des estimations selon la catégorie de taille pour les régions urbaines. Les fichiers d'enquête de microdonnées à grande diffusion ne renferment pas d'identificateurs géographiques inférieurs au niveau régional (Atlantique, Québec, Ontario, Prairies, Colombie-Britannique).

Autres variables - Le fichier de microdonnées à grande diffusion de l'enfant seulement

Les valeurs d'un certain nombre de variables ont été regroupées ou la valeur minimale et maximale des variables a été plafonnée pour des raisons de confidentialité. Voici certains exemples de ces mesures adoptées.

AGE, FAGE et SAGE

L'âge de l'enfant, du premier parent et du deuxième parent est entré dans l'un des groupes d'âges au fichier de microdonnées à grande diffusion.

FIR_EDUC et SEC_EDUC

L'année la plus élevée de scolarité du premier parent et du deuxième parent de l'enfant est inscrite dans l'une des cinq catégories au fichier de microdonnées à grande diffusion.

A1

Le nombre d'enfants de 18 ans ou moins qui habitent dans le ménage est plafonné à quatre enfants ou plus au fichier de microdonnées à grande diffusion.

Q1

La relation entre l'enfant et la personne qui donne l'information sur l'enfant est entrée en trois catégories au fichier de microdonnées à grande diffusion.

ORIGIN

Les antécédents ethniques des parents des enfants du ménage sont supprimés au fichier de microdonnées à grande diffusion. Une variable qui indique les antécédents ethniques combinés des parents des enfants (ORIGIN) est plutôt entrée au fichier de microdonnées à grande diffusion.

MAINLANG et OTHLANG

La principale langue parlée dans le ménage (MAINLANG) est entrée dans l'une des quatre catégories au fichier de microdonnées à grande diffusion. Les détails sur les autres langues parlées dans le ménage sont supprimés. Une variable qui indique si d'autres langues sont parlées ou non dans le ménage (OTHLANG) est plutôt entrée au fichier de microdonnées.

G9FGH, G9KL et G9MN

Certaines sources de revenu sont regroupées au fichier de microdonnées à grande diffusion.

8.0 Qualité des données

8.1 Taux de réponse

Le tableau qui suit renferme un résumé des taux de réponse au questionnaire de l'Enquête sur la population active (EPA) et à celui de l'Enquête sur les approches en matière de planification des études (EAPE).

Provinces	Taux de réponse des ménages pour l'EPA complète * octobre 1999	Taux de réponse des ménages pour les groupes de renouvellement de l'EPA 1, 2, 3, 5 et 6	Taux de réponse des ménages à l'EAPE **	Nombre d'enfants de 0 à 18 ans des répondants à l'EAPE	Taux de réponse des enfants à l'EAPE ***
	%	%	%		%
Terre-Neuve-et-Labrador	92,6	93,1	91,5	828	85,8
Île-du-Prince-Édouard	94,9	96,1	89,4	526	78,5
Nouvelle-Écosse	90,7	91,8	87,6	1 197	76,9
Nouveau-Brunswick	90,0	90,6	86,9	1 147	78,2
Québec	94,5	95,6	90,4	3 942	81,8
Ontario	96,1	96,9	86,0	6 161	75,7
Manitoba	94,8	96,4	86,3	1 477	76,6
Saskatchewan	94,6	95,2	85,0	1 638	74,1
Alberta	95,1	95,7	84,1	1 740	72,7
Colombie-Britannique	95,1	95,5	83,3	1 697	73,3
Canada	94,5	95,4	86,9	20 353	76,9

* Le taux de réponse à l'EPA est le nombre de ménages ayant répondu au questionnaire de l'EPA exprimé sous forme de pourcentage du nombre de ménages sélectionnés pour l'EPA.

** Le taux de réponse à l'EAPE est le nombre de ménages ayant répondu au questionnaire de l'EAPE exprimé sous forme de pourcentage du nombre de ménages qui étaient admissibles au questionnaire de l'EAPE. (Seuls les ménages de l'EPA dans les groupes de renouvellement 1, 2, 3, 5 et 6 qui avaient autorisé la communication téléphonique étaient admissibles à l'EAPE.)

*** Le taux de réponse global à l'EAPE est le nombre d'enfants pour lesquels une réponse au supplément de l'EAPE a été obtenue, exprimé en pourcentage du nombre d'enfants admissibles dans les ménages retenus pour l'EAPE. Un maximum de trois enfants par ménage pouvait être sélectionné. Les taux de réponse étaient beaucoup plus bas pour la catégorie des 16 à 18 ans, soit 69,7 %.

8.2 Erreurs relatives à l'enquête

Les estimations calculées à partir de cette enquête reposent sur un échantillon de ménages. Des estimations légèrement différentes auraient pu être obtenues si un recensement complet avait été effectué en reprenant le même questionnaire et en faisant appel aux mêmes intervieweurs, superviseurs, méthodes de traitement, etc. que ceux effectivement utilisés dans l'enquête. L'écart entre les estimations découlant de l'échantillon et celles que donnerait un dénombrement complet réalisé dans des conditions semblables est appelé erreur d'échantillonnage de l'estimation.

Des erreurs qui ne sont pas liées à l'échantillonnage peuvent se produire à presque toutes les étapes des opérations d'enquête. Les intervieweurs peuvent avoir mal compris les instructions, les enquêtés peuvent se tromper en répondant aux questions, les réponses peuvent être mal saisies sur le questionnaire et des erreurs peuvent survenir lors du traitement et de la totalisation des données. Ces erreurs sont toutes des exemples d'erreur non dues à l'échantillonnage.

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations calculées à partir de l'enquête. Toutefois, les erreurs systématiques contribuent à biaiser les estimations de l'enquête. Énormément de temps et d'efforts ont été consacrés à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été prises à chacune des étapes du cycle de collecte et de traitement des données afin de contrôler la qualité des données. Ces mesures comprenaient le recours à des intervieweurs hautement qualifiés, une formation poussée des intervieweurs concernant les procédures d'enquête et du questionnaire, l'observation des intervieweurs en vue de cerner les problèmes liés à la conception du questionnaire ou à une mauvaise compréhension des instructions, des procédures visant à s'assurer que les erreurs de saisie des données étaient réduites au minimum ainsi que des vérifications de la qualité du codage et de contrôle ayant pour but d'attester la logique du traitement.

8.2.1 Base de sondage

Comme l'EAPE était une enquête supplémentaire à l'EPA, la base de sondage de l'EPA fut utilisée. Toute non-réponse à l'EPA avait un impact sur la base de sondage de l'EAPE. La qualité des variables d'échantillon dans la base était très élevée. L'échantillon de l'EAPE était composé de cinq groupes de renouvellement de l'EPA. Les critères de sélection utilisés pour l'EAPE (tel que le groupe de renouvellement) n'étaient manquant pour aucun des enregistrements de l'EPA.

Il est à noter que la base de sondage de l'EPA exclut environ 2 % de tous les ménages dans les 10 provinces du Canada. La base de sondage de l'EAPE exclut donc la même proportion de ménages dans les mêmes régions géographiques. Il est peu probable que cette exclusion introduise un biais important dans les données de l'enquête.

Toutes les variables dans la base de sondage de l'EPA sont mises à jour tous les mois.

8.2.2 Collecte des données

La formation des intervieweurs pour l'EAPE comprenait la lecture du Manuel du surveillant, du Manuel des procédures, du Manuel de l'intervieweur et une discussion avec l'intervieweur principal des questions de l'intervieweur avant le début de l'enquête. L'information fournie aux intervieweurs comprenait une description du contexte et des objectifs de l'enquête ainsi qu'un glossaire de terminologie et une série de questions et réponses. Les intervieweurs ont recueilli l'information pour l'EAPE après avoir recueilli l'information de l'EPA.

Une lettre de présentation provenant des bureaux régionaux, a été envoyée par la poste aux répondants 10 jours avant le début de la collecte des données. L'objectif principal de la lettre était d'améliorer la qualité des données obtenues. La lettre informait les répondants que l'enquête porterait sur la collecte de données sur les épargnes pour les études postsecondaires. Les répondants pouvaient donc examiner toute l'information pertinente aux épargnes, au besoin, avant que l'on communique avec eux pour l'enquête.

8.2.3 Traitement des données

Les principales variables ont été vérifiées à 100 % pendant l'entrée des données pour en garantir la qualité. Les principales variables étaient le nombre d'enfants dans le ménage, l'âge et le sexe des enfants, l'âge de l'enfant lorsqu'on a commencé à épargner pour lui, les genres de régimes d'épargne, toutes les sommes épargnées, le revenu et la permission d'échanger les données et de nouer des liens entre elles.

Au cours du traitement des données de l'EAPE, on a découvert que 504 enregistrements ne correspondaient pas à la liste des étiquettes créées au bureau central. Il a été déterminé que la majorité de ces enregistrements supplémentaires provenaient de ménages ayant répondu à l'EPA, mais qui n'avaient pas autorisé la communication téléphonique. Les intervieweurs ont cependant pu obtenir des entrevues en personne dans ces ménages. Il a été décidé de retenir ces enregistrements.

Le traitement des données de l'EAPE a été fait par étapes, y compris la vérification, le codage, le contrôle, l'imputation, l'estimation, la confidentialité, etc. À chaque étape, une photo des fichiers de sortie est prise et il est facile de faire une vérification en comparant les fichiers de l'étape en cours avec ceux de l'étape antérieure. Cette façon de procéder a beaucoup amélioré l'étape de traitement des données.

8.2.4 Non-réponse

L'une des principales sources d'erreurs non dues à l'échantillonnage observées dans le cadre des enquêtes est l'effet de la non-réponse sur les résultats de l'enquête. L'étendue d'une non-réponse varie d'une non-réponse partielle (le fait de ne pas répondre simplement à une ou à des questions) à une non-réponse totale. S'il y a eu non-réponse totale, c'est parce que l'intervieweur a été incapable de communiquer avec le répondant, qu'aucun membre du ménage n'a pu fournir l'information demandée ou que le répondant a refusé de participer à l'enquête. Les non-réponses totales ont été traitées en ajustant le poids des ménages qui ont répondu au questionnaire d'enquête de façon à le contrebalancer pour ceux qui n'y ont pas répondu.

Dans la plupart des cas, il y a eu non-réponse partielle au questionnaire d'enquête lorsque le répondant n'a pas compris ou a mal interprété une question, a refusé d'y répondre ou ne pouvait se rappeler l'information demandée.

Dans le cas de l'EAPE de 1999, l'imputation par enregistrement donneur a permis de remplacer les données manquantes pour le revenu des ménages et six questions clés. L'imputation a servi à produire des données complètes permettant d'estimer les totaux (par exemple, le total des cotisations à un Régime enregistré d'épargne-études (REEE) en Ontario).

Les six questions clés visaient à recueillir de l'information sur la valeur actuelle de l'épargne ou sur la cotisation annuelle à cette épargne en prévision des études postsecondaires des enfants âgés de 0 à 18 ans. Les instruments d'épargne sont répartis en deux catégories : les REEEs ou les autres épargnes (c.-à-d., dépôts à terme, certificats de placement garanti (CPG), obligations d'épargne, Régimes enregistrés d'épargne-retraite (REER), fonds communs de placement). Le tableau qui suit présente les modules correspondants, les codes des questions et la description des questions.

Module	Nom de la variable	Description de la variable
G - Renseignements généraux	G10C	Revenu du ménage
B, C et D - Épargne en prévision des études postsecondaires	Q32A	Cotisations à un instrument d'épargne en 1998
	Q33A	Cotisations à un instrument d'épargne jusqu'à maintenant en 1999
	Q34A	Cotisations prévues à un instrument d'épargne pour le reste de l'année 1999
	Q35A	Valeur courante de l'épargne
	Q40A	Valeur courante de l'épargne dans un REEE
	Q43A	Valeur courante de l'épargne par d'autres personnes

Nota : Toutes les variables dérivées reliées à ces variables sont également basées sur des données imputées. Par exemple, la variable dérivée INCOMEDV du fichier de microdonnées à grande diffusion de l'enfant a été dérivée à partir de la variable imputée G10C.

Comme les six questions dépendent des questions précédentes (questions d'introduction), les valeurs manquantes pour ces dernières ont été imputées en premier. Dans les questions d'introduction, on demande si le ménage épargne (ou prévoit épargner) et, le cas échéant, si les sommes épargnées sont destinées à financer les études postsecondaires d'enfants âgés de 0 à 18 ans. Le tableau qui suit présente les modules correspondants, les codes des questions et la description des questions.

Module	Question d'introduction	Description de la question d'introduction	Question d'enchaînement
B, C et D - Planification en prévision des études postsecondaires	Q20	Les parents/tuteurs ont-ils déjà épargné en prévision des études postsecondaires de l'enfant?	Q32A, Q33A, Q34A, Q35A
	Q37A	Les parents se prévalent-ils d'un REEE pour épargner en prévision des études postsecondaires de l'enfant?	Q40A
	Q42	D'autres personnes disposent-elles d'une épargne en prévision des études postsecondaires de l'enfant?	Q43A

L'imputation a consisté à remplacer les valeurs manquantes sur le revenu du ménage, les six questions et (ou) les questions d'introduction d'un enregistrement donné (enregistrement « receveur ») par celles d'un autre enregistrement dont toutes les valeurs étaient connues et dont les caractéristiques étaient les plus « proches » (enregistrement « donneur »). Les caractéristiques de chaque enregistrement receveur ont été comparées à celles de chaque enregistrement donneur compris dans un groupe de donneurs. Lorsque la caractéristique est la même pour le receveur et un donneur, le poids (valeur) de cette caractéristique était ajouté au « score » du donneur en question. En bout de ligne, le donneur dont le score est le plus élevé est considéré comme étant le plus proche et est par conséquent choisi pour remplacer la(les) valeur(s) manquante(s) dans l'enregistrement receveur. Si plusieurs enregistrements donneurs obtiennent le score le plus élevé, l'un d'eux est choisi au hasard. Le groupe d'enregistrements

donneurs est formé de telle façon que la valeur attribuée par imputation à enregistrement receveur, conjuguées aux valeurs non imputées comprises dans cet enregistrement, ne soient pas rejetées lors des vérifications.

L'imputation par enregistrement donneur a été réalisée en trois étapes. On a d'abord imputé les données sur le revenu du ménage (G10C), en partie parce que cette variable est un facteur important dans le score de l'enregistrement donneur lors de l'imputation des valeurs pour les questions clés. On a ensuite imputé les valeurs pour les cinq variables sur l'épargne des parents (Q32A, Q33A, Q34A, Q35A et Q40A) et pour les questions d'introduction correspondantes (Q20 et Q37A). L'imputation pour ces variables a été faite simultanément pour des raisons de convergence et de cohérence. Enfin, on a imputé simultanément les valeurs pour la question sur l'épargne par d'autres (Q43A) et la question d'introduction correspondante (Q42).

Le tableau ci-dessous montre les taux d'imputation pour le revenu du ménage et les six questions.

Variable	Tous les enregistrements			Excluant les enchaînements valides		
	Imputé	Total	Taux (%)	Imputé	Total	Taux (%)
G10C	1 525	11 544	13	1 525	11 544	13
Q32A	3 025	20 353	15	2 976	8 496	35
Q33A	3 388	20 353	17	3 339	8 496	39
Q34A	4 062	20 353	20	4 013	8 496	47
Q35A	2 367	20 353	12	2 318	8 496	27
Q40A	1 144	20 353	6	749	3 202	23
Q43A	3 398	20 353	17	1 648	2 565	64

Il convient de noter que la variable G10C est connue au niveau du ménage tandis que les autres variables relatives à l'épargne sont connues au niveau de l'enfant. Les taux présentés sont calculés en fonction de tous les enregistrements et des enregistrements pour lesquels la valeur ne correspond pas à un enchaînement valide.

Le processus d'imputation de l'EAPE a donné de bons résultats. Dans l'ensemble, il a permis de remplir les réponses incomplètes à partir des renseignements fournis par d'autres répondants ayant des caractéristiques semblables ou identiques, et d'accroître le nombre d'unités pouvant servir aux analyses des chercheurs.

8.2.5 Mesure de l'erreur d'échantillonnage

Puisqu'il est inévitable que des estimations établies à partir d'une enquête-échantillon (ou par sondage) soient sujettes à une erreur d'échantillonnage, une saine pratique de la statistique exige que les chercheurs fournissent aux utilisateurs une certaine indication de l'importance de cette erreur d'échantillonnage. La présente section de la documentation renferme un aperçu des mesures de l'erreur d'échantillonnage dont Statistique Canada se sert couramment et dont le Bureau conseille vivement aux utilisateurs qui produisent des estimations à partir de ce fichier de microdonnées à employer également.

La base pour mesurer l'importance potentielle des erreurs d'échantillonnage est l'erreur-type des estimations calculées à partir des résultats d'une enquête.

En raison, cependant, de la diversité des estimations pouvant être produites à partir d'une enquête, l'erreur-type d'une estimation est habituellement exprimée en fonction de l'estimation à laquelle elle se rapporte. La mesure résultante, appelée coefficient de variation (CV) d'une estimation, s'obtient en divisant l'erreur-type de l'estimation par l'estimation elle-même et s'exprime en pourcentage de l'estimation.

Par exemple, supposons que, d'après les résultats de l'enquête, l'on estime que 15,6 % des enfants ont un Régime enregistré d'épargne-études (REEE)⁵ au Canada et l'on constate que l'erreur-type de cette estimation est de 0,0048. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left(\frac{0,0048}{0,156} \right) \times 100 \% = 3,1 \%$$

De plus amples renseignements sur le calcul du coefficient de variation, se trouvent au chapitre 10.0.

⁵ Argent que les membres du ménage ont versé dans un REEE pour l'enfant.

9.0 Lignes directrices pour la totalisation, l'analyse et la diffusion de données

Ce chapitre de la documentation renferme un aperçu des lignes directrices que doivent respecter les utilisateurs qui totalisent, analysent, publient ou autrement diffusent des données calculées à partir des fichiers de microdonnées de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs de microdonnées de produire les mêmes chiffres que ceux produits par Statistique Canada, tout en étant en mesure d'obtenir des chiffres actuellement inédits de façon conforme à ces lignes directrices établies.

9.1 Lignes directrices pour l'arrondissement d'estimations

Afin que les estimations qui sont destinées à la publication ou à tout autre forme de diffusion qui sont calculées à partir de ces fichiers de microdonnées correspondent à celles produites par Statistique Canada, nous conseillons vivement aux utilisateurs à respecter les lignes directrices qui suivent en ce qui concerne l'arrondissement de telles estimations :

- a) Les estimations dans le corps principal d'un tableau statistique doivent être arrondies à la centaine près à l'aide de la technique d'arrondissement normale. Selon cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, le dernier chiffre à conserver est augmenté de 1. Par exemple, selon la technique d'arrondissement normale à la centaine près, si les deux derniers chiffres se situent entre 00 et 49, ils sont remplacés par 00 et le chiffre précédent (le chiffre des centaines) reste inchangé. Si les derniers chiffres se situent entre 50 et 99, ils sont remplacés par 00 et le chiffre précédent est augmenté de 1.
- b) Les totaux partiels marginaux et des totaux marginaux des tableaux statistiques doivent être calculés à partir de leurs composantes non arrondies correspondantes, puis ensuite être arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement normale.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir de composantes non arrondies (c'est-à-dire des numérateurs et/ou des dénominateurs), puis être arrondis à leur tour à une décimale à l'aide de la technique d'arrondissement normale. Dans le cas d'un arrondissement normal à un seul chiffre, si le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, le dernier chiffre à conserver est augmenté de 1.
- d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs composantes non arrondies correspondantes, puis être arrondis à leur tour à la centaine près (ou à la décimale près) à l'aide de la technique d'arrondissement normale.
- e) Dans les cas, où, en raison de limitations d'ordre technique ou de toutes autres limites, une technique d'arrondissement autre que la technique normale est utilisée produisant des estimations à être publiées ou autrement diffusées différentes des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces différences dans le ou les documents à publier ou à diffuser.
- f) En aucun cas, les utilisateurs ne doivent publier ou autrement diffuser des estimations non arrondies. Des estimations non arrondies laissent entendre qu'elles sont plus précises qu'elles le sont en réalité.

9.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation

Le plan d'échantillonnage utilisé pour l'Enquête sur les approches en matière de planification des études (EAPE) n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer le poids d'échantillonnage approprié.

Si les poids appropriés ne sont pas utilisés, les estimations calculées à partir des fichiers de microdonnées ne peuvent être considérées comme représentatives de la population visée par l'enquête et ne correspondront pas à celles produites par Statistique Canada.

Les utilisateurs devraient également prendre note que certains progiciels pourraient peut-être ne pas permettre la production d'estimations correspondant exactement à celles qu'offre Statistique Canada, en raison du mode de traitement du champ du poids par ces progiciels.

9.3 Définitions de types d'estimations : catégoriques et quantitatives

Le fichier de microdonnées à grande diffusion de l'EAPE a été structuré de façon à ce que l'enfant soit l'unité d'analyse. Le poids de chaque enregistrement est un coefficient de pondération « enfant ». Les estimations des parents, des familles ou des ménages ne peuvent être calculées à partir du fichier de microdonnées de l'EAPE.

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'EAPE, il est utile de décrire les deux principaux types d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites à partir du fichier de microdonnées créé pour l'EAPE.

9.3.1 Estimations catégoriques

Les estimations catégoriques sont des estimations du nombre ou du pourcentage de membres de la population visée par l'enquête possédant certaines caractéristiques ou faisant partie d'une catégorie définie. Le nombre d'enfants qui ont des épargnes pour les études postsecondaires ou la proportion d'enfants qui ont recours à des Régimes enregistrés d'épargne-études constituent des exemples de telles estimations. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être désignée une estimation d'un agrégat.

Exemples de questions catégoriques :

Q : Les parents de (nom de l'enfant) ont-ils aménagé un espace à la maison pour qu'il(elle) puisse faire ses travaux scolaires ?

R : Oui / Non

Q : Dans la liste suivante, indiquez quel aspect financier est le plus prioritaire.

R : Budget quotidien / Épargnes pour les études postsecondaires / Épargnes pour la retraite / Autres épargnes

9.3.2 Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures d'une tendance centrale de quantités reposant sur

certaines ou sur tous les membres de la population visée par l'enquête. Elles comprennent aussi expressément des estimations de la forme \hat{X} / \hat{Y} où \hat{X} est une estimation de la quantité totale pour la population visée par l'enquête et \hat{Y} , est une estimation du nombre de personnes dans la population visée par l'enquête qui contribuent à cette quantité totale.

Un exemple d'estimation quantitative est le nombre moyen d'épargnes pour les études réalisées pour les enfants de moins de six ans. Le numérateur est une estimation de la somme totale des épargnes pour les enfants de moins de six ans et son dénominateur est le nombre d'enfants de moins de six ans pour lesquels des épargnes ont été déclarées.

Exemples de questions quantitatives :

Q : Dans l'éventualité où (nom de l'enfant) continuerait à vivre à la maison, à combien ses parents/tuteurs estiment-ils les coûts globaux de ses études ?

R : |_|_|_| |_|_|_| \$

Q : Combien d'argent a été économisé pour les études postsecondaires de (nom de l'enfant) en 1998 ?

R : |_|_|_| |_|_|_| \$

9.3.3 Totalisation d'estimations catégoriques

On peut obtenir des estimations du nombre de gens possédant une certaine caractéristique à partir du fichier de microdonnées en additionnant les poids finals de tous les enregistrements possédant la ou les caractéristiques qui nous intéressent. On obtient des proportions et des rapports de la forme \hat{X} / \hat{Y} en :

- a) additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le numérateur (\hat{X}),
- b) additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le dénominateur (\hat{Y}), puis en
- c) divisant l'estimation a) par celle de b) (\hat{X} / \hat{Y}).

9.3.4 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable qui nous intéresse par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements qui nous intéressent. Pour obtenir, par exemple, une estimation de la somme totale d'argent épargné pour les études postsecondaires des enfants dont les parents espèrent qu'ils iront à l'université, multipliez la valeur déclarée à la question Q35A (somme totale épargnée) par le poids final de l'enregistrement, puis additionnez cette valeur pour tous les enregistrements où la variable Q3 = 4 (université).

Pour obtenir une moyenne pondérée de la forme \hat{X} / \hat{Y} , le numérateur (\hat{X}) est calculé comme une estimation quantitative et le dénominateur (\hat{Y}) est calculé comme une estimation catégorique. Pour estimer, par exemple, la somme moyenne d'argent épargné

pour les études postsecondaires des enfants dont les parents espèrent qu'ils iront à l'université,

- a) estimez le nombre total d'argent (\hat{X}) tel qu'il est décrit ci-dessus,
- b) estimez le nombre de personnes (\hat{Y}) incluses dans cette catégorie en additionnant les poids finals de tous les enregistrements où la variable Q3 = 4, puis
- c) divisez l'estimation a) par l'estimation b) (\hat{X} / \hat{Y}).

9.4 Lignes directrices pour l'analyse statistique

L'EAPE repose sur un plan d'échantillonnage complexe comportant une stratification, de multiples étapes de sélection ainsi que des probabilités inégales de sélection des répondants. L'utilisation des données provenant d'enquêtes aussi complexes présente des problèmes pour les analystes, parce que le plan d'enquête et les probabilités de sélection influent sur les procédures d'estimation et de calcul de la variance qui devraient être utilisées. Il faut utiliser les poids de l'enquête pour que les estimations et les analyses des données de l'enquête soient exemptes de biais.

Bien que de nombreuses procédures d'analyse que l'on trouve à l'intérieur de progiciels statistiques permettent d'utiliser des poids, la stratification ou la définition du poids inclus dans ces procédures peut différer de ce qui convient dans le contexte d'une enquête-échantillon, de telle sorte que dans bien des cas les estimations produites au moyen de ces progiciels sont correctes, mais que les variances calculées sont piètres. Les variances approximatives pour des estimations simples comme des totaux, des proportions et des rapports (pour des variables qualitatives) peuvent être calculées à partir des tables de variabilité d'échantillonnage approximative qui accompagnent les données.

Pour d'autres techniques d'analyse (de régression linéaire, de régression logistique et de l'analyse de variance, par exemple), il existe une méthode qui peut rendre les variances calculées par l'application des progiciels normalisés plus significatives, en intégrant les probabilités inégales de sélection. L'application de cette méthode entraîne une remise à l'échelle des poids de façon à ce que le poids moyen soit de 1.

Supposons, par exemple, qu'il faut effectuer l'analyse de tous les répondants de sexe masculin. Les étapes à suivre pour remettre à l'échelle les poids sont les suivantes :

- 1) sélectionner tous les répondants du fichier qui ont déclaré SEXE = masculin;
- 2) calculer le poids MOYEN pour ces enregistrements en additionnant les poids originaux des enfants établis à partir du fichier de microdonnées pour ces enregistrements puis diviser cette somme par le nombre de répondants ayant déclaré SEXE = masculin;
- 3) pour chacun de ces répondants, calculer un poids REMIS À L'ÉCHELLE égal au poids original de l'enfant divisé par le poids MOYEN;
- 4) effectuer l'analyse portant sur ces enfants en utilisant le poids REMIS À L'ÉCHELLE.

Parce qu'on ne tient toujours compte ni de la stratification ni des grappes du plan d'échantillonnage, les estimations des variances calculées avec cette méthode risquent cependant d'être des sous-estimations.

Il faut connaître les détails du plan d'enquête pour calculer des estimations des variances plus précises. De tels détails ne peuvent être fournis dans le fichier de microdonnées en raisons de la

confidentialité. Statistique Canada peut, contre remboursement des frais, calculer des variances qui tiennent compte du plan complet d'échantillonnage pour beaucoup de statistiques.

9.5 Lignes directrices pour la diffusion de coefficients de variation

Avant de diffuser et/ou de publier toutes estimations établies à partir de l'EAPE, les utilisateurs devraient premièrement déterminer le niveau de qualité de cette estimation. Les niveaux de qualité sont *acceptable*, *médiocre* et *inacceptable*. Les erreurs d'échantillonnage et non dues à l'échantillonnage, dont il a été question au chapitre 8.0, influencent la qualité des données. Aux fins du présent document, cependant, on ne déterminera le niveau de qualité d'une estimation qu'à partir d'une erreur d'échantillonnage dont rend compte le coefficient de variation indiqué à l'intérieur du tableau qui figure ci-dessous. Les utilisateurs devraient néanmoins s'assurer de lire le chapitre 8.0 pour être plus pleinement informés des caractéristiques relatives à la qualité de ces données.

On devrait premièrement déterminer le nombre de répondants retenus pour le calcul de l'estimation. Si ce nombre est inférieur à 30, il faudrait considérer l'estimation pondérée comme étant de qualité inacceptable.

Pour les estimations pondérées fondées sur les tailles d'échantillon de 30 ou plus, les utilisateurs devraient déterminer le coefficient de variation de l'estimation et suivre les lignes directrices relatives au niveau de qualité qui figurent ci-dessous. Celles-ci devraient être appliquées, pour la détermination du niveau de qualité d'une estimation, aux estimations pondérées arrondies.

On peut considérer qu'il est possible de divulguer toutes les estimations. Celles d'un niveau de qualité médiocre ou inacceptable doivent cependant être accompagnées d'une mise en garde pour avertir les utilisateurs subséquents.

Lignes directrices relatives au niveau de qualité de l'estimation

Niveau de qualité de l'estimation	Lignes directrices
1) Acceptable	<p>Les estimations proviennent d'une taille d'échantillon de 30 ou plus, et présentent de faibles coefficients de variation, de l'ordre de 0,0 à 16,5 %.</p> <p>Aucune mise en garde n'est requise.</p>
2) Médiocre	<p>Les estimations proviennent d'une taille d'échantillon de 30 ou plus, et présentent des coefficients de variation élevés, de l'ordre de 16,6 à 33,3 %.</p> <p>Ces estimations devraient être signalées par la lettre M (ou un quelconque identificateur similaire). Elles devraient être accompagnées d'une mise en garde avertissant les utilisateurs subséquents des niveaux élevés d'erreur associés aux estimations.</p>
3) Inacceptable	<p>Les estimations proviennent d'une taille d'échantillon inférieure à 30, ou présentent des coefficients de variation très élevés, supérieurs à 33,3 %.</p> <p>Statistique Canada recommande de ne pas diffuser d'estimations de qualité inacceptable. Si un utilisateur choisit cependant de le faire, ces estimations devraient alors être signalées à l'aide de la lettre I (ou d'un quelconque identificateur similaire) et devraient être accompagnées de la mise en garde suivante :</p> <p>« Nous informons l'utilisateur que ces estimations (désignées avec la lettre I) ne respectent pas les normes de qualité de Statistique Canada. Les conclusions qui reposeront sur ces données ne seront ni fiables et seront très probablement invalides. »</p>

9.6 Seuils pour la diffusion des estimations pour l'Enquête sur les approches en matière de planification des études

Le tableau ci-dessous fournit une indication de la précision des estimations des prévisions démographiques ainsi que les seuils de diffusion associés aux trois niveaux de qualité de l'estimation présentés à la section précédente. Ces seuils proviennent des tables de coefficients de variation (CV) dont il sera question au chapitre 10.0.

Par exemple, d'après le tableau, la qualité d'une estimation pondérée de 5 000 personnes possédant une caractéristique donnée à Terre-Neuve-et-Labrador est médiocre.

Veillez noter que ces seuils de diffusion correspondent aux estimations de chiffres de population seulement. Dans le cas d'estimations de rapports, les utilisateurs ne devraient pas utiliser la valeur du numérateur (ni le dénominateur) afin de trouver le niveau de qualité de l'estimation correspondant. La règle 4 à la section 10.1 ainsi que l'exemple 4 à la section 10.1.1 expliquent la bonne procédure à suivre dans le cas d'un rapport.

Provinces et régions	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Terre-Neuve-et-Labrador	11 000 et plus	3 000 à < 11 000	moins de 3 000
Île-du-Prince-Édouard	5 000 et plus	1 500 à < 5 000	moins de 1 500
Nouvelle-Écosse	15 500 et plus	4 000 à < 15 500	moins de 4 000
Nouveau-Brunswick	10 000 et plus	2 500 à < 10 000	moins de 2 500
Québec	50 000 et plus	12 500 à < 50 000	moins de 12 500
Ontario	49 500 et plus	12 500 à < 49 500	moins de 12 500
Manitoba	15 500 et plus	4 000 à < 15 500	moins de 4 000
Saskatchewan	13 000 et plus	3 500 à < 13 000	moins de 3 500
Alberta	33 000 et plus	8 500 à < 33 000	moins de 8 500
Colombie-Britannique	42 000 et plus	10 500 à < 42 000	moins de 10 500
Provinces de l'Atlantique	12 000 et plus	3 000 à < 12 000	moins de 3 000
Provinces des Prairies	25 000 et plus	6 500 à < 25 000	moins de 6 500
Canada	40 500 et plus	10 000 à < 40 500	moins de 10 000

10.0 Tables de variabilité d'échantillonnage approximative

Afin de fournir des coefficients de variation (CV) qui pourraient s'appliquer à une gamme étendue d'estimations catégoriques produites à partir de ce fichier de microdonnées et auxquels il serait facilement possible pour l'utilisateur d'avoir accès, un ensemble de tables de variabilité d'échantillonnage approximative a été produit. Ces tables de CV permettent à l'utilisateur d'obtenir un coefficient de variation approximatif fondé sur la taille de l'estimation calculée à partir des données d'enquête.

Les coefficients de variation sont calculés à l'aide de la formule de la variance pour un échantillonnage aléatoire simple et en y incorporant un facteur qui reflète la nature du plan d'échantillonnage, qui est à plusieurs degrés et qui prévoit la formation de grappes. Ce facteur, appelé l'effet du plan, a été déterminé en calculant premièrement les effets du plan pour une gamme étendue de caractéristiques, puis en choisissant parmi ceux-ci une valeur modérée (habituellement le 75^e percentile) à utiliser à l'intérieur de ces tables de CV qui s'appliqueraient ensuite à l'ensemble entier de caractéristiques.

Le tableau ci-dessous indique la valeur modérée des effets du plan, ainsi que les tailles de l'échantillon et les chiffres de population selon la province pour les enfants âgés de 0 à 18 ans qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative de l'Enquête sur les approches en matière de planification des études (EAPE).

Provinces et régions	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	2,14	828	129 085
Île-du-Prince-Édouard	2,30	526	35 859
Nouvelle-Écosse	2,42	1 197	221 251
Nouveau-Brunswick	1,89	1 147	175 739
Québec	3,26	3 942	1 695 246
Ontario	2,96	6 161	2 863 607
Manitoba	2,39	1 477	279 577
Saskatchewan	2,29	1 638	266 933
Alberta	2,09	1 740	786 608
Colombie-Britannique	2,17	1 697	935 429
Provinces de l'Atlantique	2,22	3 698	561 934
Provinces des Prairies	2,55	4 855	1 333 118
Canada	3,07	20 353	7 389 334

Tous les coefficients de variation inclus dans les tables de variabilité d'échantillonnage approximative sont approximatifs et donc non officiels. Des estimations de la variance réelle pour des variables précises peuvent être obtenues auprès de Statistique Canada, contre remboursement des frais. Étant donné que le CV approximatif est une estimation prudente, l'utilisation de la variance réelle estimée pourrait permettre aux utilisateurs de passer l'estimation d'un niveau de qualité à un autre. Par exemple, une estimation *médiocre* pourrait devenir *acceptable* si elle était fondée sur le calcul du CV exact.

Rappelez-vous que : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée est très probablement inacceptable et Statistique Canada recommande de ne pas diffuser une telle estimation, quelle que soit la valeur du coefficient de variation.

10.1 Comment utiliser les tables de coefficients de variation pour des estimations catégoriques

Les règles qui suivent devraient permettre à l'utilisateur de déterminer les coefficients de variation approximatifs à partir des tables de variabilité d'échantillonnage approximative pour des estimations du nombre, de la proportion ou du pourcentage de membres de la population visée par l'enquête possédant une certaine caractéristique et pour des rapports et des différences entre de telles estimations.

Règle 1 : Estimations du nombre d'enfants possédant une caractéristique donnée (agrégats)

Le coefficient de variation dépend uniquement de la taille de l'estimation elle-même. Dans la table de variabilité d'échantillonnage approximative pour la région géographique appropriée, repérez le nombre estimé dans la colonne la plus à gauche (intitulée « Numérateur du pourcentage ») et suivez les astérisques (le cas échéant) jusqu'au premier chiffre rencontré. Ce chiffre est le coefficient de variation approximatif.

Règle 2 : Estimations de proportions ou de pourcentages d'enfants possédant une caractéristique donnée

Le coefficient de variation d'une proportion estimée ou d'un pourcentage estimé dépend à la fois de la taille de la proportion ou du pourcentage et de la taille du total sur lequel la proportion ou le pourcentage repose. Les proportions estimées ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque la proportion ou que le pourcentage repose sur un sous-groupe de la population. La proportion, par exemple, d'enfants qui ont des épargnes pour les études postsecondaires, est plus fiable que le nombre estimé d'enfants qui ont des épargnes pour les études postsecondaires. (Remarquez que dans les tables la valeur des coefficients de variation diminue lorsqu'on les lit de gauche à droite.)

Lorsque la proportion ou que le pourcentage repose sur la population totale de la région géographique visée par la table, le CV de la proportion ou du pourcentage est le même que le CV du numérateur de la proportion ou du pourcentage. Dans ce cas, la règle 1 peut être appliquée.

Lorsque la proportion ou que le pourcentage repose sur un sous-ensemble de la population totale (p. ex., comme ses membres d'un sexe ou d'un groupe d'âge particulier), on devrait faire référence à la proportion ou au pourcentage (dans le haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table). L'intersection de la rangée et de la colonne appropriées donne le coefficient de variation.

Règle 3 : Estimations de différences entre des agrégats ou des pourcentages

L'erreur-type d'une différence entre deux estimations est approximativement égale à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément. C'est-à-dire que l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 est l'estimation 1, \hat{X}_2 est l'estimation 2 et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement. Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d}$. Cette

formule est exacte pour la différence entre des caractéristiques distinctes et non corrélées, mais n'est autrement qu'approximative.

Règle 4 : Estimations de rapports

Dans le cas où le numérateur est un sous-ensemble du dénominateur, le rapport devrait être converti en un pourcentage et la règle 2 appliquée. Cela s'appliquerait, par exemple, au cas où le dénominateur est le nombre d'enfants et le numérateur, le nombre d'enfants qui ont des épargnes pour les études postsecondaires.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre d'enfants qui ont des épargnes et dont les parents espèrent qu'ils iront à l'université comparativement au nombre d'enfants qui ont des épargnes et dont les parents espèrent qu'ils iront au collège, l'erreur-type du rapport des estimations est approximativement égal à la racine carrée de la somme des carrés de chaque coefficient de variation considéré séparément multipliée par \hat{R} . C'est-à-dire que l'erreur-type d'un rapport ($\hat{R} = \hat{X}_1 / \hat{X}_2$) est :

$$\sigma_{\hat{R}} = \hat{R} \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et de \hat{X}_2 respectivement. Le coefficient de variation de \hat{R} est donné par $\sigma_{\hat{R}} / \hat{R}$. La formule tendra à surestimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés positivement et à la sous-estimer si \hat{X}_1 et \hat{X}_2 sont corrélés négativement.

Règle 5 : Estimations de différences entre des rapports

Dans ce cas, les règles 3 et 4 sont combinées. On détermine premièrement les CV pour les deux rapports à l'aide de la règle 4, puis on trouve le CV de leur différence au moyen de la règle 3.

10.1.1 Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques

Les exemples ci-dessous utilisent des données du fichier de l'EAPE et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter.

Exemple 1 : Estimations du nombre d'enfants possédant une caractéristique donnée (agrégats)

Supposons qu'un utilisateur estime que 3 008 755 le nombre d'enfants qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation des enfants pour le CANADA.
- 2) L'agrégat estimé (3 008 755) ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, c'est-à-dire 3 000 000.
- 3) On trouve le coefficient de variation pour un agrégat estimé en se reportant à la première entrée autre que des astérisques sur cette rangée, c'est-à-dire 1,4 %.

- 4) Le coefficient de variation approximatif de l'estimation est donc 1,4 %. Le résultat selon lequel il y avait 3 008 755 (à être arrondi d'après les lignes directrices tel qu'il est indiqué à la section 9.1) d'enfants qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires, peut être publié sans réserve.

Exemple 2 : Estimations de proportions ou de pourcentages d'enfants possédant une caractéristique donnée

Supposons qu'un utilisateur estime que les parents/tuteurs de 2 396 350 / 3 008 755 = 79,6 % des d'enfants qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires, espèrent qu'ils iront à l'université. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation des enfants pour le CANADA.
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les enfants qui ont des épargnes pour les études postsecondaires), il faut utiliser à la fois le pourcentage (79,6 %) et la portion numérateur du pourcentage (2 396 350) pour déterminer le coefficient de variation.
- 3) Le numérateur, 2 396 350, ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, soit 2 000 000. De même, l'estimation du pourcentage ne figure sous aucun des titres de colonne; il faut donc utiliser la proportion qui s'en rapproche le plus, soit 70,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 1,3 %, est le coefficient de variation à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 1,3 %. Le résultat selon lequel les parents/tuteurs de 79,6 % des enfants qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires, espèrent qu'ils iront à l'université, peut être publié sans réserve.

Exemple 3 : Estimations de différences entre des agrégats ou des pourcentages

Supposons qu'un utilisateur estime que les parents/tuteurs de 1 224 140 / 1 493 156 = 82,0 % des filles qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires, espèrent qu'elles iront à l'université tandis que les parents/tuteurs de 1 172 211 / 1 515 599 = 77,3 % des garçons qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires, espèrent qu'ils iront à l'université. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) L'utilisation de la table des coefficients de variation des enfants pour le CANADA de la même façon que celle décrite dans l'exemple 2, donne un CV de l'estimation pour les filles de 1,1 % et un CV de l'estimation pour les garçons de 1,8 %.
- 2) En utilisant la règle 3, l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 est l'estimation 1 (filles), \hat{X}_2 est l'estimation 2 (garçons), et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et de \hat{X}_2 respectivement.

C'est-à-dire que l'erreur-type de la différence $\hat{d} = 0,820 - 0,773 = 0,047$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0,820)(0,011)]^2 + [(0,773)(0,018)]^2} \\ &= \sqrt{(0,000081) + (0,000194)} \\ &= 0,017\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,017 / 0,047 = 0,362$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 36,2 %. Cette estimation est considérée inacceptable et Statistique Canada recommande de ne pas publier cette estimation. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre I (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime que les parents/tuteurs de 1 224 140 de filles qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires espèrent qu'elles iront à l'université, tandis que les parents de 1 172 211 de garçons qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires espèrent qu'ils iront à l'université. L'utilisateur est intéressé à comparer l'estimation des filles à celle des garçons sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation (\hat{X}_1) est le nombre de filles qui ont des épargnes pour les études et dont les parents ou tuteurs espèrent qu'elles iront à l'université. Le dénominateur de l'estimation (\hat{X}_2) est le nombre de garçons qui ont des épargnes pour les études et dont les parents ou tuteurs espèrent qu'ils iront à l'université.
- 2) Reportez-vous à la table des coefficients de variation des enfants pour le CANADA.
- 3) Le numérateur de cette estimation de rapport est 1 224 140. Le chiffre qui s'en rapproche le plus est 1 000 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 3,1 %.
- 4) Le dénominateur de cette estimation de rapport est 1 172 211. Le chiffre qui s'en rapproche le plus est 1 000 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 3,1 %.
- 5) Le coefficient de variation approximatif de l'estimation du rapport est donné par la règle 4, qui est :

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement.

C'est-à-dire que :

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0,031)^2 + (0,031)^2} \\ &= \sqrt{(0,000961) + (0,000961)} \\ &= 0,044 \end{aligned}$$

- 6) Le rapport obtenu entre les filles et les garçons qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires et dont les parents/tuteurs espèrent qu'ils iront à l'université est 1 224 140 / 1 172 211, c'est-à-dire 1,04:1 (à être arrondi d'après les lignes directrices tel qu'il est indiqué à la section 9.1). Le coefficient de variation de cette estimation est 4,4 %, ce qui fait qu'on peut la diffuser sans réserve.

10.2 Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance

Bien que les coefficients de variation sont beaucoup utilisés, l'intervalle de confiance d'une estimation est une mesure plus intuitivement significative de l'erreur d'échantillonnage. Un intervalle de confiance constitue une déclaration du niveau de confiance selon laquelle la valeur vraie pour la population se situe à l'intérieur d'une gamme précisée de valeurs. Par exemple, un intervalle de confiance de 95 % peut être décrit comme suit :

Si l'échantillonnage de la population est répète indéfiniment, chaque échantillon menant à un nouvel intervalle de confiance pour une estimation, l'intervalle englobera alors dans 95 % des échantillons la valeur vraie de la population.

En utilisant l'erreur-type d'une estimation, des intervalles de confiance pour des estimations peuvent être obtenues en partant de l'hypothèse qu'aux termes d'un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique donnée de la population se répartiront normalement autour de la valeur vraie de la population. Selon cette hypothèse, il y a environ 68 chances sur 100 que l'écart entre une estimation de l'échantillon et la valeur vraie pour la population soit inférieur à une erreur-type, environ 95 chances sur 100 que l'écart soit inférieur à deux erreurs-types et environ 99 chances sur 100 que l'écart soit inférieur à trois erreurs-types. Ces différents degrés de confiance sont désignés sous le nom de niveaux de confiance.

Des intervalles de confiance pour une estimation \hat{X} sont généralement exprimés sous forme de deux chiffres, un inférieur et un supérieur à l'estimation, comme étant $(\hat{X} - k, \hat{X} + k)$, où k est déterminé suivant le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

Des intervalles de confiance pour une estimation peuvent être calculés directement à partir des tables de variabilité d'échantillonnage approximative, en déterminant d'abord à

partir de la table appropriée le coefficient de variation de l'estimation \hat{X} , puis en utilisant la formule suivante pour le convertir à un intervalle de confiance ($IC_{\hat{x}}$) :

$$IC_{\hat{x}} = (\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}})$$

où $\alpha_{\hat{x}}$ est le coefficient de variation déterminé de \hat{X} , et

- $t = 1$ si l'on désire un intervalle de confiance de 68 %;
- $t = 1,6$ si l'on désire un intervalle de confiance de 90 %;
- $t = 2$ si l'on désire un intervalle de confiance de 95 %;
- $t = 2,6$ si l'on désire un intervalle de confiance de 99 %.

Nota : Les lignes directrices pour la diffusion des estimations s'appliquent également aux intervalles de confiance. S'il est impossible, par exemple, de diffuser une estimation, on ne peut alors pas non plus communiquer un intervalle de confiance.

10.2.1 Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance

Un intervalle de confiance de 95 % pour la proportion estimée d'enfants qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires et dont les parents/tuteurs espèrent qu'ils iront à l'université (d'après l'exemple 2 à la section 10.1.1) serait calculé comme suit :

$$\hat{X} = 79,6 \% \text{ (ou exprimé sous forme de proportion } 0,796)$$

$$t = 2$$

$\alpha_{\hat{x}} = 1,3 \%$ (0,013 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.

$$IC_{\hat{x}} = \{0,796 - (2) (0,796) (0,013), 0,796 + (2) (0,796) (0,013)\}$$

$$IC_{\hat{x}} = \{0,796 - 0,021, 0,796 + 0,021\}$$

$$IC_{\hat{x}} = \{0,775, 0,817\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 77,5 % et 81,7 % des enfants qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires, les parents/tuteurs espèrent qu'ils iront à l'université.

10.3 Comment utiliser les tables de coefficients de variation pour effectuer un test t

Des erreurs-types peuvent aussi être utilisés pour effectuer des tests d'hypothèses, une procédure destinée à distinguer des paramètres d'une population à l'aide d'estimations d'un échantillon. Ces estimations peuvent être des chiffres, des moyennes, des pourcentages, des rapports, etc. Les tests peuvent être effectués à divers niveaux de signification, où un niveau de signification est la probabilité de conclure que les caractéristiques sont différentes quand, en fait, elles sont identiques.

Supposons que \hat{X}_1 et \hat{X}_2 sont des estimations d'un échantillon pour deux caractéristiques qui nous intéressent. Supposons également que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $\sigma_{\hat{d}}$.

Si $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ se situe entre -2 et 2, aucune conclusion à propos de la différence entre les

caractéristiques n'est alors justifiée au niveau de signification de 5 %. Si, cependant, ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0,05. C'est-à-dire que la différence entre les estimations est significative.

10.3.1 Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion de filles qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires, dont les parents/tuteurs espèrent qu'elles iront à l'université et la proportion de garçons qui, grâce à des membres de leur ménage, ont des épargnes en vue de leurs études postsecondaires, dont les parents/tuteurs espèrent qu'ils iront à l'université. D'après l'exemple 3 à la section 10.1.1, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,017. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,820 - 0,773}{0,017} = \frac{0,047}{0,017} = 2,76$$

Puisque $t = 2,76$ est plus grand que 2, il faut en conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 0,05.

10.4 Coefficients de variation pour des estimations quantitatives

Il faudrait produire des tables spéciales afin de déterminer l'erreur d'échantillonnage d'estimations quantitatives, ce qui n'a pas été fait, parce que la plupart des variables pour l'EAPE sont principalement de nature catégorique.

En général cependant, le coefficient de variation d'un total quantitatif sera supérieur au coefficient de variation de l'estimation de la catégorie correspondante (c'est-à-dire l'estimation du nombre de personnes retenues dans l'estimation quantitative). S'il est impossible de diffuser l'estimation de la catégorie correspondante, on ne pourra pas non plus communiquer l'estimation quantitative. Par exemple, le coefficient de variation de la somme totale versée dans des Régimes enregistrés d'épargne-études serait supérieur au coefficient de variation de la proportion correspondante d'enfants qui ont des épargnes dans des Régimes enregistrés d'épargne-études. Si, par conséquent, le coefficient de variation de la proportion est inacceptable (rendant la proportion non diffusable), il en sera de même du coefficient de variation de l'estimation quantitative correspondante (rendant cette estimation quantitative non diffusable).

Des coefficient de variation de telles estimations peuvent être calculés, au besoin, pour une estimation précise à l'aide d'une technique appelée pseudo-répétition, ce qui veut dire diviser les enregistrements inclus dans les fichiers de microdonnées en sous-groupes (ou répétitions) et déterminer la variation à l'intérieur de l'estimation de répétition en répétition. Les utilisateurs qui

désirent calculer des coefficients de variation pour des estimations quantitatives peuvent communiquer avec Statistique Canada afin d'obtenir des conseils sur l'allocation d'enregistrements à des répétitions appropriées et sur les formules à employer à l'intérieur de ces calculs.

10.5 Tables des coefficients de variation

Voir le fichier EAPE1999_Maître_Enfant_CVTabSF.pdf pour les tables de coefficients de variation a utilisées avec le fichier de microdonnées à grande diffusion (FMGD) de l'enfant et le fichier maître de l'enfant.

11.0 Pondération

Puisque l'Enquête sur les approches en matière de planification des études (EAPE) faisait appel à un sous-échantillon de l'échantillon de l'Enquête sur la population active (EPA), le calcul des poids pour les enregistrements des données de l'enquête est clairement lié à la procédure de pondération utilisée aux fins de l'EPA. La procédure de pondération employée pour l'EPA est décrite brièvement ci-dessous.

11.1 Procédures de pondération pour l'Enquête sur la population active

Dans le cas de l'EPA, le poids final rattaché à chaque enregistrement est le produit des facteurs suivants : le poids de base, le sous-poids d'une grappe, le poids de stabilisation, le facteur compensatoire pour les non-réponses et le facteur d'ajustement du rapport province-âge-sexe et du niveau infraprovincial. Chacun de ces facteurs est décrit ci-dessous.

Poids de base

Dans le cas d'un échantillon probabiliste, le plan d'échantillonnage lui-même détermine les poids qui doivent être utilisés pour produire des estimations non biaisées de la population. Il faut pondérer chaque enregistrement à l'aide de l'inverse de la probabilité de sélectionner la personne à qui l'enregistrement renvoie. Dans le cas d'un échantillon aléatoire simple de 2 %, comme cette probabilité serait de 0,02 pour chaque personne, il faut pondérer les enregistrements à l'aide de $1 / 0,02 = 50$. Étant donné la complexité du plan de l'EPA, les logements situés dans différentes régions auront des poids de base différents. Parce que toutes les personnes admissibles habitant un logement sont interviewées (directement ou par personne interposée), cette probabilité est essentiellement la même que la probabilité de sélection du logement.

Sous-poids d'une grappe

La délimitation des grappes est telle que le nombre de logements inclus dans l'échantillon augmente très légèrement en cas de croissance modérée du parc immobilier. On peut tolérer une croissance importante à l'intérieur d'une grappe isolée avant que l'échantillon additionnel ne pose un problème sur le plan de la collecte des données sur le terrain. S'il y a cependant croissance dans plus d'une grappe incluse à l'intérieur de la tâche d'un intervieweur, l'effet cumulatif de toutes les augmentations peut créer un problème sur le plan de la charge de travail. Dans le cas des grappes dont la croissance est importante, on utilise un sous-échantillon pour que les tâches des intervieweurs demeurent gérables. Le sous-poids d'une grappe représente l'inverse de ce rapport de sous-échantillonnage dans le cas des grappes pour lesquelles il y a eu sous-échantillonnage.

Poids de stabilisation

On a aussi recours à la stabilisation d'un échantillon pour s'attaquer aux problèmes soulevés par la croissance de la taille de ce dernier. Le sous-échantillonnage d'une grappe s'attaquait à une croissance isolée dans des régions relativement petites, tandis que la stabilisation d'un échantillon s'attaque à la croissance lente d'un échantillon au fil du temps qui est le résultat d'un taux fixe d'échantillonnage parallèlement à une augmentation générale de la taille, ou de l'effectif, de la population. La stabilisation d'un échantillon est la suppression aléatoire de logements de l'échantillon, ce qui vise à maintenir la taille de ce dernier à son niveau désiré. On ajuste le poids de base à l'aide du rapport de la taille de l'échantillon, qui repose sur le taux fixe d'échantillonnage, à la taille de l'échantillon désirée. On appelle ce facteur d'ajustement le poids de stabilisation. L'ajustement se fait à l'intérieur de secteurs de stabilisation définis comme étant des logements appartenant à la même région économique de l'assurance-emploi et au même groupe de renouvellement.

Non-réponse

Dans le cas de certains types de non-réponses (comme les ménages temporairement absents ou les refus), les données de l'interview menée le mois précédent auprès d'un ménage, le cas échéant, sont réutilisées pour ce ménage comme données du mois courant.

Dans d'autres cas, on contrebalance les poids des non-réponses en accroissant proportionnellement ceux des ménages ayant répondu au questionnaire. On augmente le poids de chaque enregistrement d'un ménage ayant répondu au questionnaire à l'aide du rapport du nombre de ménages qui auraient dû être interviewés divisé par le nombre de ceux qui l'ont réellement été. On effectue séparément cet ajustement pour les secteurs de non-réponses, qui sont définis par la région économique de l'assurance-emploi, le type de secteur et le groupe de renouvellement. L'ajustement repose sur l'hypothèse voulant que les ménages qui ont été interviewés représentent les caractéristiques de ceux qui auraient dû être interviewés se trouvant à l'intérieur d'un secteur de non-réponse.

Sous-poids de l'Enquête sur la population active

On appelle le sous-poids de l'EPA le produit des facteurs de pondération décrits précédemment. Tous les membres du même ménage échantillonné ont le même sous-poids.

Ajustements au niveau infraprovincial et au rapport province-âge-sexe

On peut utiliser le sous-poids afin de calculer une estimation valable de toute caractéristique pour laquelle on collecte des données au moyen de l'EPA. Cependant, ces estimations seront fondées sur une base contenant des renseignements qui peuvent être périmés depuis plusieurs années et qui ne sont donc pas représentatifs de la population actuelle. Au moyen de renseignements complémentaires plus récents sur la population cible, les poids d'échantillonnage sont ajustés en vue d'améliorer la précision des estimations et la représentativité de l'échantillon de la population actuelle.

On dispose mensuellement d'estimations indépendantes pour divers groupes d'âge-sexe selon la province. Ce sont des projections démographiques fondées sur les données du recensement, les enregistrements des naissances et des décès et les estimations de la migration, les plus récents. À la dernière étape, on utilise ces renseignements auxiliaires pour convertir le sous-poids en poids final, ce qui se fait à l'aide d'une méthode de calibration. Cette méthode assure que les poids finals qu'elle produit équivalent aux projections du recensement pour les variables auxiliaires, c'est-à-dire des totaux pour divers groupes d'âge-sexe, de régions économiques, de régions métropolitaines de recensement, de groupes de renouvellement, de ménages et de la taille de la famille économique. On corrige également les poids de manière à ce que la somme des estimations de la branche d'activité et de la main-d'œuvre du mois précédent, dérivées de l'échantillon du mois en cours, correspondent aux estimations correspondantes de l'échantillon du mois précédent. Il s'agit de l'estimation composite. On applique la méthode de la régression généralisée à l'ensemble de la correction.

Habituellement, on ne tient pas compte de ce poids définitif dans la détermination du poids d'une enquête supplémentaire de l'EPA. On a plutôt recours à la sous-pondération comme l'expliquent les paragraphes qui suivent.

11.2 Procédures de pondération pour l'Enquête sur les approches en matière de planification des études

Les principes qui sous-tendent le calcul des poids pour l'Enquête sur les approches en matière de planification des études sont identiques à ceux établis aux fins de l'EPA. Toutefois, d'autres

ajustements sont apportés aux sous-poids de l'EPA afin de calculer un poids final pour les différents enregistrements inclus dans le fichier de microdonnées de l'EAPE.

- 1) Un ajustement qui tient compte de l'utilisation d'un sous-échantillon représentant les cinq sixièmes, plutôt que de l'échantillon complet de l'EPA.
- 2) Un ajustement qui tient compte des non-réponses de ménages additionnelles au questionnaire de l'enquête supplémentaire, c'est-à-dire des non-réponses au questionnaire de l'EAPE attribuables à des gens qui ont répondu au questionnaire de l'EPA ou pour qui des données de l'EPA du mois précédent ont été reportées. La procédure est similaire à l'ajustement des poids des non-réponses au questionnaire de l'EPA, mais les groupements reposent sur des variables différentes.
- 3) Un rajustement qui tient compte des projections distinctes du nombre de ménages selon la province et la taille des ménages (un, deux et trois personnes ou plus), après avoir apporté les ajustements ci-dessus.

Voici d'où sont tirés les poids pour les enfants :

- 4) Un ajustement apporté au poids ajusté pour les non-réponses établies à l'ajustement 2) ci-dessus, afin de tenir compte du sous-échantillonnage des enfants des ménages qui comptent plus de trois enfants.
- 5) Un rajustement pour tenir compte des projections distinctes du nombre de personnes selon la province, le groupe d'âge et le sexe. Voici les groupes d'âges utilisés : de 0 à 5, de 6 à 12, de 13 à 15 et de 16 à 18.

On tient compte des ajustements 1) et 2) en multipliant le sous-poids de l'EPA pour chaque enregistrement affichant une réponse au questionnaire de l'EAPE par :

$$= \left(\frac{6}{5} \right) * \left(\frac{\text{la somme des sous-poids de l'EPA de chaque ménage dans les groupes de renouvellement 1, 2, 3, 5 et 6 qui ont répondu à l'EPA}}{\text{la somme des sous-poids de l'EPA de chaque ménage qui ont répondu à l'EAPE}} \right)$$

pour obtenir un sous-poids (WEIGHT1) ajusté pour les non-réponses des ménages.

Des estimations distinctes du nombre de ménages qui comptent un, deux et trois personnes ou plus selon la province sont disponibles chaque mois. Il s'agit de projections démographiques déterminées selon les données du Recensement de 1996, les enregistrements des naissances et des décès et les estimations de la migration. L'étalonnage a permis d'obtenir le poids final des ménages et de garantir que les poids finals obtenus équivalent aux projections du recensement pour ces variables auxiliaires. Cette méthode accentue la fiabilité des estimations que l'on peut tirer de l'Enquête sur les approches en matière de planification des études. L'ajustement 3) a été calculé en multipliant le WEIGHT1 pour chaque répondant à l'EAPE par :

$$= \frac{\text{le nombre prévu de ménages se trouvant } t \text{ dans la province et le groupe de la taille du ménage}}{\text{la somme de WEIGHT 1 pour les ménages qui ont répondu à l'EAPE se trouvant } t \text{ dans la province et le groupe de la taille des ménages}}$$

Le poids obtenu est le poids final des ménages qui apparaît au fichier maître des ménages de l'Enquête sur les approches en matière de planification des études que conserve Statistique Canada. Il est à noter que le poids des ménages n'est pas disponible au fichier de microdonnées à grande diffusion.

On procède aux ajustements 4) et 5) pour calculer un poids pour chaque enfant sur lequel il y a une réponse au questionnaire de l'EAPE. L'ajustement 4) est apporté en multipliant le WEIGHT1 par :

$$= \frac{\text{le nombre d'enfants de 18 ans et moins dans le ménage}}{\text{minimum nombre d'enfants de 18 ans et moins dans le ménage, maximum de 3}}$$

pour obtenir un poids pour enfant WEIGHT3 ajusté du sous-échantillon.

Des estimations distinctes sont aussi disponibles chaque mois selon l'âge et le sexe et la province. On a procédé à l'étalonnage pour obtenir le poids final pour enfant, afin de garantir que les poids finals produits équivalent aux projections du recensement selon la province, le sexe et les groupes d'âges. Les groupes d'âges étaient les 0 à 5, 6 à 12, 13 à 15 et 16 à 18.

Voilà qui améliore la fiabilité des estimations qui peuvent être tirées de l'Enquête sur les approches en matière de planification des études. L'ajustement 5) a été calculé en multipliant le WEIGHT1 pour chaque répondant à l'EAPE par :

$$= \frac{\text{le nombre prévu d'enfants se trouvant dans la province, le sexe et le groupe d'âge}}{\text{la somme de WEIGHT 3 pour les enfants des répondants à l'EAPE se trouvant dans la province, le sexe et le groupe d'âge}}$$

Le poids obtenu (FINWT) est le poids final pour enfants entré au fichier maître et au fichier de microdonnées à grande diffusion de l'enfant de l'Enquête sur les approches en matière de planification des études.

N'oubliez pas que la variable FINWT (12,4) au fichier maître de l'enfant et au fichier de microdonnées à grande diffusion de l'enfant a une décimale physique de quatre chiffres de la droite (9999999.9999).

12.0 Questionnaires

12.1 Le questionnaire de l'Enquête sur la population active

Le questionnaire de l'Enquête sur la population active (EPA_QuestF.pdf) sert à recueillir des données au sujet de l'activité actuelle et de la dernière activité sur le marché du travail de tous les membres d'un ménage âgés de 15 ans ou plus. Il comprend des questions sur les heures de travail, l'ancienneté dans l'emploi, le type de travail, la raison des heures perdues ou d'absence, la recherche d'emploi réalisée, la disponibilité pour le travail et la fréquentation scolaire.

12.2 Le questionnaire de l'Enquête sur les approches en matière de planification des études

Le questionnaire de l'Enquête sur les approches en matière de planification des études (EAPE) a servi à recueillir des données pour l'enquête supplémentaire en octobre 1999. Le fichier EAPE1999_QuestF.pdf renferme le questionnaire français.

L'Enquête sur les approches en matière de planification des études de 1999 comprend au total 171 questions réparties en sept sections (A à G). La première section (A) comporte une question de sélection visant à faire en sorte que les répondants se limitent aux questions qui correspondent à leur situation. Les trois sections qui suivent (B, C et D) portent sur les expériences scolaires et les épargnes en vue des études de trois enfants du ménage. La section E porte sur les épargnes en vue des études de tout autre enfant âgé de 18 ans ou moins faisant partie du ménage. La section F comporte des questions au sujet des épargnes en vue des études pour les enfants qui ne vivent pas dans le ménage. La dernière section (G) comprend des questions de nature plus générale, en vue de dresser le profil de certaines caractéristiques du ménage.

Section A : Questions de sélection

Les questions de cette section visent à déterminer si des enfants âgés de 18 ans ou moins vivent habituellement dans le ménage. À partir des réponses à ces questions, le répondant peut se rendre à la section appropriée du questionnaire. Lorsque le ménage compte plus de trois enfants âgés de 18 ans ou moins, trois d'entre eux sont choisis au hasard.

Section B : Enfant 1

Cette section comporte des questions au sujet du premier enfant du ménage âgé de 18 ans ou moins. Les questions portent sur les expériences scolaires de l'enfant, au cours de la dernière année, sur les attentes des parents/tuteurs quant aux coûts des études postsecondaires de cet enfant et sur les plans pour défrayer ces coûts. Si des épargnes ont été effectuées pour l'enfant, des questions portent sur les montants épargnés, la façon dont les épargnes sont investies et la proportion des épargnes investies dans des Régimes enregistrés d'épargne-études.

Section C : Enfant 2

Cette section comporte les mêmes questions que la section B pour le deuxième enfant du ménage âgé de 18 ans ou moins.

Section D : Enfant 3

Cette section comporte les mêmes questions que la section B pour le troisième enfant du ménage âgé de 18 ans ou moins.

Section E : Autres enfants du ménage : épargnes en vue des études postsecondaires

Cette section comporte des questions au sujet des épargnes en vue des études pour tout autre enfant âgé de 18 ans ou moins qui vit dans le ménage. Les questions servent à recueillir des données au sujet de la nature et du montant des épargnes.

Section F : Enfants hors du ménage

Cette section comporte des questions au sujet des épargnes en vue des études pour des enfants qui ne vivent pas dans le ménage. Les questions servent à recueillir des données au sujet de la nature et du montant des épargnes.

Section G : Renseignements généraux

Les questions de cette section servent à recueillir des données au sujet des ressources dont le ménage dispose, de l'origine ethnique, de la langue, des priorités financières et du revenu.

13.0 Clichés d'enregistrements à valeurs univariées

Voir le fichier EAPE1999_FMGD_Enfant_LvCds.pdf pour le cliché d'enregistrement à chiffres univariés pour le fichier de microdonnées à grande diffusion (FMGD) de l'enfant.

Voir le fichier EAPE1999_Maître_Enfant_LvCds.pdf pour le cliché d'enregistrement à chiffres univariés pour le fichier maître de l'enfant.

13.1 Unité d'analyse

Selon la structure du fichier de l'Enquête sur les approches en matière de planification des études (EAPE), l'enfant est l'unité d'analyse. Le fichier contient un enregistrement par enfant. Le poids tiré de chaque enregistrement est une pondération « enfant ». Après pondération, les enregistrements de ce fichier donnent le nombre total d'enfants âgés de 0 à 18 ans au Canada. Le fichier de microdonnées de l'EAPE ne permet pas de faire des estimations sur les parents, les familles ou les ménages.

Le fichier de microdonnées étant structuré au niveau de l'enfant, il exclut les ménages qui n'avaient pas d'enfants qui résidaient dans le ménage au moment de l'enquête. Les microdonnées ne comprennent pas non plus les variables de la section E (les épargnes des autres enfants du ménage lorsqu'il y a plus de trois enfants) ou de la section F (les épargnes des enfants qui habitent à l'extérieur du ménage du répondant). Certaines variables dérivées au fichier de microdonnées sont créées à l'aide des données de ces sections, c'est-à-dire les variables dérivées qui sont axées sur les épargnes du ménage. Il est possible d'obtenir davantage d'information détaillée sur les ménages sans enfants et les sections E et F en consultant les totalisations personnalisées et disponibles à la Sous-section des services à la clientèle de la Division des enquêtes spéciales (voir chapitre 1.0).

13.2 Disponibilité de l'information sur les parents

L'information sociodémographique et sur la population active au sujet de la mère et/ou du père de l'enfant est disponible pour la majorité des enfants. Dans deux cas cependant, les données ne sont pas disponibles pour un parent ou les deux. Premièrement, si un parent n'habite pas dans le même ménage que l'enfant, les données ne sont pas disponibles pour ce parent. Deuxièmement, si une relation parent-enfant ne peut être tirée des données de l'Enquête sur la population active, les données ne sont pas disponibles pour les parents.

La relation parent-enfant est tirée de l'information obtenue au cours de l'Enquête sur la population active au sujet de la relation entre les gens dans une famille économique. Une famille économique est un groupe de deux personnes ou plus qui habitent dans le même logement et qui sont apparentées par le sang, le mariage, le droit commun ou l'adoption. La relation entre les membres de la famille économique est déterminée en fonction du répondant à l'Enquête sur la population active ou la « personne de référence ». Nous sommes donc informés de la relation de la personne de référence avec chaque membre de la famille économique, mais nous ignorons les relations des membres de la famille entre eux. Prenons, par exemple, le cas d'un enfant qui habite avec sa mère et sa tante. Si la mère est la personne de référence, sa relation avec l'enfant est inscrite dans le questionnaire de l'Enquête sur la population active comme « fils - fille ». Si la tante est la personne de référence, cependant, sa relation avec l'enfant est inscrite comme suit : « frère ou soeur ou autre parent » et sa relation avec la mère est inscrite comme suit : « frère ou soeur ou autre parent ». La relation entre la mère et l'enfant est inconnue. Dans le deuxième cas, il n'y aura pas d'information sur les parents de l'enfant au fichier de microdonnées à grande diffusion parce que la relation parent-enfant n'a pu être déterminée.

Une autre variable tirée du point de vue du répondant à l'Enquête sur la population active est EFAMTYPE qui décrit le type de famille économique. Le type de famille économique est déterminé selon le point de vue de la personne de référence de l'Enquête sur la population active, et non du point de vue de l'enfant. Si l'enfant habite avec sa mère et son père et son père est la personne de référence de l'EPA, le type de famille économique serait « époux-épouse ». Si l'enfant habite cependant avec sa mère et ses grands-parents, et sa grand-mère est la personne de référence de l'EPA, le type de famille économique serait « époux-épouse ». Dans ce cas cependant, la désignation « époux-épouse » ferait référence aux grands-parents de l'enfant, et non à ses parents.

13.3 Convention de désignation des variables

L'information sur les enfants des ménages a été obtenue à l'aide des Sections B, C et D du questionnaire de l'Enquête sur les approches en matière de planification des études. La numérotation du questionnaire est établie selon la section. Dans la section B, par exemple, l'information obtenue porte sur le premier enfant du ménage et la première question est donc numérotée B1.

Il y a un enregistrement par enfant au fichier de microdonnées. Le nom des variables pour les sections B, C et D est devenu Q1 (équivalant à B1, C1 ou D1), Q2 (équivalant à B2, C2 ou D2), etc.