

TABLE DES MATIÈRES

	Page
1. Introduction	6
2. Renseignements de base	7
3. Objectifs	7
4. Contenu de l'enquête	8
5. Plan d'échantillonnage	10
5.1 Répartition de l'échantillon	10
5.2 Méthode d'exclusion	12
5.3 Prélèvement de l'échantillon	12
5.4 Intégration de l'ENSP et de l'Enquête longitudinale nationale sur les enfants (ELNE)	14
5.5 Plan d'échantillonnage au Québec	15
6. Collecte des données	16
6.1 Conception du questionnaire et méthode de collecte des données	16
6.2 Essais	17
6.3 Interviews	18
6.4 Surveillance et contrôle	18
6.5 Non-réponse à l'ENSP	19
6.6 Suivi des cas de non-réponse	19
7. Traitement des données	20
7.1 Saisie des données	20
7.2 Vérification	20
7.3 Codage	20
7.4 Création de variables calculées	20
7.5 Facteurs de pondération	21
7.6 Suppression de renseignements confidentiels	21
8. Qualité des données	22
8.1 Taux de réponse	22
8.2 Erreurs relatives à l'enquête	24
9. Lignes directrices pour la totalisation, l'analyse et la diffusion	26
9.1 Lignes directrices pour l'arrondissement	26

9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	27
9.2.1	Définitions des genres d'estimations : de type nominal par opposition à quantitatives	27
9.2.2	Totalisation d'estimations de type nominal	28
9.2.3	Totalisation d'estimations quantitatives	29
9.3	Lignes directrices pour l'analyse statistique	29
9.4	Lignes directrices pour la diffusion	30
10.	Tables de variabilité d'échantillonnage approximative	32
10.1	Comment utiliser les tables de c.v. pour les estimations de type nominal	34
10.2	Exemples d'utilisation des tables de c.v. pour des estimations de type nominal	37
10.3	Comment utiliser les tables de c.v. pour calculer les limites de confiance	40
10.4	Exemple d'utilisation de tables de c.v. pour obtenir des limites de confiance	41
10.5	Comment utiliser les tables de c.v. pour effectuer un test	41
10.6	Exemple d'utilisation des tables de c.v. pour effectuer un test t	42
10.7	Coefficients de variation d'estimations quantitatives	42
10.8	Seuils pour la diffusion des estimations relatives à l'ENSP	44
11.	Facteurs de pondération	47
11.1	Méthode de pondération pour les provinces à part le Québec	47
11.1.1	Poids de base de l'EPA	47
11.1.2	Facteurs de pondération de base de la composition aléatoire	48
11.1.3	Corrections supplémentaires apportées aux facteurs de pondération de base	48
11.1.4	Corrections supplémentaires pour les membres du ménage	52
11.1.5	Corrections supplémentaires pour les membres du panel	54
11.2	Méthode de pondération pour le Québec	59
11.2.1	Facteurs de pondération de l'ESS	59
11.2.2	Facteurs de pondération de base pour les logements de l'ENSP	59
11.2.3	Corrections supplémentaires aux facteurs de pondération de base	61
11.2.4	Corrections supplémentaires pour les membres du ménage	63
11.2.5	Corrections supplémentaires pour les membres du panel	64
Annexe A: Questionnaire		
Annexe B: Cliché d'enregistrement - fichier de micro-données - Général		
Annexe C: Cliché d'enregistrement - fichier de micro-données - Santé		
Annexe D: Dictionnaire des données - fichier de micro-données - Général		
Annexe E: Dictionnaire des données - fichier de micro-données - Santé		

Annexe F: Variables dérivées

Annexe G: Tableaux C.V.

Tableaux C.V., Canada selon le groupe d'âge, fichier de micro-données - Général

Tableaux C.V. selon les provinces et le Canada, fichier de micro-données - Général

Tableaux C.V., Canada selon le groupe d'âge, fichier de micro-données - Santé

Tableaux C.V., selon les provinces et le Canada, fichier de micro-données - Santé

GUIDE DU FICHIER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

1. Introduction

L'Enquête nationale sur la santé de la population (ENSP) vise à recueillir de l'information sur la santé de la population canadienne. Le premier cycle de collecte de données, d'une durée de 12 mois, a débuté en 1994; par la suite, d'autres collectes de données auront lieu tous les deux ans. L'enquête servira à recueillir non seulement des données transversales, mais aussi des données auprès d'un échantillon permanent d'individus à intervalles de deux ans.

La population cible de l'ENSP comprend les résidents des ménages de toutes les provinces, à l'exclusion de la population des réserves indiennes, des bases des Forces armées canadiennes et de certaines régions éloignées du Québec et de l'Ontario. Des enquêtes distinctes ont été menées au Yukon, dans les Territoires du Nord-Ouest et dans les établissements de santé (auprès des résidents à long terme d'hôpitaux et d'établissements de soins institutionnels); les résultats de ces enquêtes seront présentés plus tard.

L'Enquête nationale sur la santé de la population (ENSP) a été menée par Statistique Canada en 1994-1995. Ce guide a pour but de faciliter la manipulation du fichier de microdonnées des résultats de l'enquête.

Pour toute question au sujet de l'ensemble de données ou de son utilisation, s'adresser à :

Jeanine Bustros
Division des statistiques sur la santé
Statistique Canada
ENSP, section L
20^e étage, édifice R.-H.-Coats
Parc Tunney
Ottawa (Ontario)
K1A 0T6
(613) 951-3285
(613) 951-4198 télécopieur

Pour toute question de nature technique s'adresser à : Maryanne Kirkpatrick (613)-951-1137

2. Renseignements de base

À l'automne de 1991, le Conseil national d'information sur la santé (CNIS) a recommandé la réalisation d'une enquête permanente nationale sur la santé de la population. Cette recommandation découlait des pressions économiques et fiscales exercées sur le régime de santé et du besoin correspondant d'information pouvant servir à améliorer l'état de santé de la population canadienne. Les sources existantes de données sur la santé sont insuffisantes pour dresser un tableau complet de l'état de santé de la population et de la multitude de facteurs qui ont un effet sur la santé.

Dès le mois d'avril 1992, Statistique Canada a obtenu des fonds pour élaborer l'Enquête nationale sur la santé de la population. L'enquête a été conçue de manière à être souple et à produire des données valables, fiables et actuelles. Il fallait aussi que l'enquête puisse être adaptée aux besoins, aux politiques et aux intérêts changeants.

3. Objectifs

Les objectifs de l'ENSP sont les suivants :

- ˘ faciliter l'élaboration de politiques gouvernementales en fournissant des données sur l'état de santé de la population, une ventilation de ces données ainsi qu'un aperçu des tendances observées;
- ˘ fournir des données en vue de la réalisation d'études analytiques qui aideront à comprendre les facteurs déterminants de la santé;
- ˘ recueillir des données sur les corrélations entre la santé et les facteurs économiques, sociaux, démographiques, professionnels et environnementaux;
- ˘ éclaircir la relation entre l'état de santé et l'utilisation des services de santé, y compris les services parallèles et les services traditionnels;
- ˘ fournir des renseignements sur un panel de personnes qui seront suivies dans le temps pour refléter la dynamique de la santé et de la maladie;
- ˘ donner aux provinces, aux territoires et à d'autres clients une capacité d'enquête sur la santé qui leur permettra de compléter le contenu ou l'échantillon;
- ˘ permettre de faire des liens entre les données de l'enquête et des données administratives courantes, telles que les statistiques de l'état civil, les mesures environnementales, les variables liées aux collectivités, ainsi que l'utilisation des services de santé.

4. Contenu de l'enquête

Ces objectifs représentent seulement l'orientation générale de l'ENSP, notamment en ce qui concerne le type de renseignements à recueillir. Par conséquent, le contenu de l'enquête a été déterminé d'après les critères suivants :

- 1) Les renseignements doivent être pertinents par rapport aux objectifs des provinces et des territoires en matière de santé et faciliter le contrôle de ces objectifs. Dans les cas où il n'existe pas d'objectifs en matière de santé, par exemple au niveau national, le contenu de l'enquête peut être déterminé à partir des politiques et des programmes.
- 2) Les renseignements ne doivent pas faire double emploi avec d'autres sources d'information.
- 3) Pour favoriser une meilleure compréhension de la santé et de ses facteurs déterminants, l'enquête doit fournir de nouveaux renseignements dans des domaines qui n'ont pas encore été étudiés de manière approfondie.
- 4) L'enquête doit être axée sur des comportements ou des conditions pouvant faire l'objet de prévention, de traitements ou d'interventions.
- 5) L'enquête doit recueillir des renseignements sur les conditions qui causent le plus de souffrance ou qui coûtent le plus cher aux personnes touchées, à l'ensemble de la population ou au régime de santé.
- 6) L'enquête doit recueillir des renseignements sur les facteurs liés à une bonne santé et non seulement sur ceux liés à la maladie.

Un certain nombre de renseignements sont recueillis auprès de tous les membres de chaque ménage; de plus, une personne âgée de 12 ans et plus est choisie au hasard dans chaque ménage pour une interview plus approfondie. Selon les critères énoncés ci-dessus, le questionnaire porte sur l'état de santé, l'utilisation des services de santé, les facteurs de risque et la situation démographique et socio-économique. Par exemple, l'état de santé est mesuré à l'aide de questions sur la perception personnelle de la santé, de la capacité fonctionnelle, des états chroniques et des limitations d'activité. L'utilisation des services de santé est mesurée au moyen de questions sur les visites faites à des professionnels de la santé, les soins hospitaliers et l'usage de médicaments. Les questions sur les facteurs de risque à l'égard du comportement portent sur l'usage du tabac, la consommation d'alcool et l'activité physique. De plus, la première enquête contenait des questions spéciales sur les facteurs psychosociaux qui peuvent influencer la santé, par exemple le stress, l'estime de soi et le soutien social. Les questions sur les caractéristiques démographiques et socio-économiques portent sur l'âge, le sexe, la scolarité, l'origine ethnique, le revenu du ménage et la situation professionnelle. L'annexe A contient la liste des questions de l'enquête.

GUIDE DU FICHIER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

5. Plan d'échantillonnage

La population cible de l'ENSP comprend les résidents des ménages de toutes les provinces, à l'exclusion de la population des réserves indiennes, des bases des Forces armées canadiennes et de certaines régions éloignées du Québec et de l'Ontario.

Plan d'échantillonnage du volet des ménages

Le plan d'échantillonnage du volet des ménages a été établi selon quatre critères :

- la taille visée des échantillons nationaux et provinciaux/territoriaux;
- la décision de choisir un membre par ménage pour constituer le panel longitudinal;
- le choix de l'Enquête sur la population active (EPA) comme véhicule pour sélectionner l'échantillon; et
- la décision d'intégrer l'ENSP et l'Enquête longitudinale nationale sur les enfants (ELNE).

Les trois premiers facteurs ont déterminé respectivement la répartition de l'échantillon, l'application d'une technique pour accroître la représentativité (la «méthode d'exclusion», décrite plus loin) et la sélection des échantillons provinciaux à l'extérieur du Québec.

5.1 Répartition de l'échantillon

Le budget de l'ENSP prévoyait un échantillon d'une taille de 19 600 ménages. Par ailleurs, les représentants nationaux et provinciaux ont convenu qu'il fallait un minimum de 1 200 ménages par province. Sous réserve de cette restriction, la taille des échantillons provinciaux a été établie à l'aide d'une méthode de répartition bien connue qui met en équilibre les exigences de fiabilité aux niveaux national et régional (Kish, 1988). Selon cette méthode, l'échantillon a été réparti proportionnellement parmi $\sqrt{(0.804W_h^2 + 1/12^2)}$, où W_h représente la proportion des ménages dans la province ($h, h=1, \dots, 12$) d'après le recensement de 1991. Cette répartition a permis de déterminer la taille de l'échantillon de base de chaque province. Quatre provinces ont choisi d'acheter des unités d'échantillonnage supplémentaires pour augmenter l'échantillon qui leur avait été alloué.

Dans chaque province, l'échantillon a été réparti initialement en proportion de la taille de la population. Les unités d'échantillonnage supplémentaires achetées par certaines provinces et l'utilisation d'une méthode d'exclusion, décrite plus loin, ont eu un effet sur la répartition infraprovinciale de l'échantillon. L'Ontario et le Manitoba ont acheté des unités d'échantillonnage supplémentaires pour satisfaire à des exigences minimales dans leurs régions sanitaires, tandis que le Nouveau-Brunswick et la Colombie-Britannique ont acheté des unités d'échantillonnage supplémentaires pour certaines régions seulement. En Colombie-Britannique, presque toutes les unités d'échantillonnage supplémentaires ont été contactées au moyen d'interviews téléphoniques, à partir d'un échantillon de numéros de téléphone établi par composition aléatoire (CA). Selon la méthode d'exclusion, la taille

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

des échantillons a été élargie en fonction du nombre de ménages que l'on s'attendait de rejeter à la sélection.

Le tableau 1 ci-dessous montre la taille prévue des échantillons selon la province. Les chiffres représentent le nombre de logements privés occupés qui étaient admissibles, avant la non-réponse (estimée à près de 10 %).

TABLEAU 1 : TAILLE DE L'ÉCHANTILLON DE L'ENSP

Taille de l'échantillon (ménages)

Province	Répartition initiale	Unités achetées	À interviewer	Rejet à la sélection	Total
T.-N.	1 220	-	1 221	171	1 392
Î.-P.-É.	1 201	-	1 199	223	1 422
N.-É.	1 270	-	1 270	246	1 516
N.-B.	1 243	180	1 423	234	1 657
Qué.	3 584	-	3 479	-	3 479
Ont.	4 817	2 183	7 001	1 021	8 022
Man.	1 307	493	1 800	324	2 124
Sask.	1 287	-	1 288	257	1 545
Alb.	1 674	-	1 674	305	1 979
C.-B. (1)	1 996	61	2 057	448	2 505
C.-B. (2)	-	788	788	-	788
TOTAL	19 599	3 705	23 200	3 229	26 429

(1) Sauf la portion CA. (2) Portion CA.

5.2 Méthode d'exclusion

Le contenu de l'enquête vise principalement un membre de chaque ménage échantillonné qui est choisi au hasard pour faire partie du panel longitudinal. Plus un ménage est grand, moins il y a de chances qu'un membre particulier soit choisi; par conséquent, les personnes appartenant à des ménages de grande taille, notamment les parents et leurs enfants, sont sous-représentées dans le panel, tandis que les personnes appartenant à des ménages de petite taille, en général les personnes célibataires et les personnes âgées, sont surreprésentées.

La méthode d'exclusion a été adoptée pour accroître la représentativité des parents et des jeunes dans le panel. Une partie de l'échantillon est soumise à une sélection préliminaire. Une fois la liste des membres du ménage dressée, les ménages n'ayant aucun membre âgé de moins de 25 ans sont exclus de l'enquête. Afin de maintenir la taille voulue des échantillons, on ajoute au nombre de ménages visités dans chaque province le nombre prévu de ménages qui sont exclus de cette façon.

La méthode d'exclusion des ménages n'ayant aucun membre âgé de moins de 25 ans a été adoptée parce qu'elle était plus efficace que d'autres méthodes qui ont été envisagées. Pour des raisons financières et opérationnelles, le pourcentage de ménages rejetés à la sélection a généralement été limité à 25-30 % en Ontario. Il était à 37,5-40 % dans les régions urbaines des autres provinces et à 25-30 % dans toutes les régions rurales. Étant donné que les strates d'appartements avaient une forte concentration de petits ménages, on a réduit la taille de ces échantillons plutôt que d'y appliquer la méthode d'exclusion. De plus, la méthode d'exclusion n'a pas été appliquée dans les régions éloignées, en raison des coûts élevés pour contacter les ménages à ces endroits. Finalement, l'utilisation de la méthode d'exclusion a été limitée dans les régions où le nombre d'unités d'échantillonnage achetées était élevé.

5.3 Prélèvement de l'échantillon

Un plan d'échantillonnage stratifié à deux degrés a été utilisé pour le volet des ménages de l'ENSP. Le premier degré consiste à former des strates homogènes et à prélever des échantillons indépendants de grappes dans chaque strate. Le deuxième degré consiste à dresser des listes de ménages pour chaque grappe et à sélectionner des logements, ou ménages, à partir de ces listes.

Dans chaque province, sauf au Québec, l'ENSP a fait appel à la méthode d'échantillonnage polyvalente mise au point pour le remaniement de l'Enquête sur la population active (EPA). Les échantillons de logements répartis en grappes qui sont constitués selon cette méthode peuvent servir à des enquêtes-ménages générales, ce qui réduit considérablement les frais de listage et de collecte de données.

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

Le plan de base de l'EPA est constitué d'un échantillon stratifié à plusieurs degrés composé de logements sélectionnés parmi les grappes. Chaque province est divisée en trois types de régions (grands centres urbains, villes urbaines et régions rurales), à partir desquelles sont formées des strates géographiques et/ou socio-économiques distinctes. Dans la plupart des strates, on choisit six grappes, généralement des secteurs de dénombrement (SD), selon une méthode d'échantillonnage avec probabilité proportionnelle à la taille (PPT). Dans quelques cas où la densité de la population est faible, on ajoute un degré supplémentaire en sélectionnant d'abord deux ou trois unités primaires d'échantillonnage, en les divisant en grappes et en prélevant un échantillon de six grappes à l'intérieur de chacune. Dans tout le plan d'échantillonnage, on utilise six grappes afin de pouvoir renouveler chaque mois le sixième de l'échantillon aux fins de l'EPA.

L'échantillon de logements est obtenu une fois que les opérations de listage des grappes sont terminées. Étant donné que les taux d'échantillonnage sont déterminés au préalable, il y a souvent un écart entre la taille prévue de l'échantillon et le nombre d'unités obtenu. Le rendement excessif de l'échantillon est corrigé en supprimant une partie des unités sélectionnées à l'origine. Cette opération, qui se fait habituellement à un niveau agrégé, est appelée stabilisation de l'échantillon. De plus, on augmente la taille de l'échantillon afin de représenter les logements plutôt que les ménages, étant donné qu'environ 15 % des logements sont habituellement vacants ou hors du champ de l'enquête.

Le plan d'échantillonnage est établi de manière à obtenir environ 60 000 ménages. Pour les enquêtes qui exigent des échantillons de moindre taille, on «réserve» habituellement de un à six renouvellements par province, un renouvellement correspondant au sixième de l'échantillon total. On procède à la stabilisation de l'échantillon pour le maintenir au niveau voulu, notamment lorsqu'on réserve deux renouvellements mais que la taille de l'échantillon requise représente seulement 1,5 renouvellement.

En raison des besoins particuliers de l'ENSP, deux modifications ont été apportées à cette stratégie d'échantillonnage. Le nombre de renouvellements «réservés» a été déterminé au niveau des strates plutôt qu'au niveau provincial, afin de répondre aux exigences relatives à la répartition infraprovinciale de l'échantillon. Il fallait aussi que le nombre de grappes sélectionnées par strate soit un multiple de quatre, aux fins de l'estimation de la variance et de la représentativité saisonnière (les strates pouvaient ainsi être réparties en deux ou plusieurs échantillons indépendants de quatre grappes chacun - soit un par période de collecte). Étant donné que l'ENSP ne nécessitait généralement que de deux à six grappes par strate de l'EPA, des strates similaires de l'EPA ont été groupées de manière à former de plus grandes strates de l'ENSP comportant le nombre requis de grappes de l'échantillon.

Par suite de ces modifications, l'échantillon de grappes de l'ENSP peut être considéré comme un échantillon stratifié répété dont les strates sont constituées de groupes de

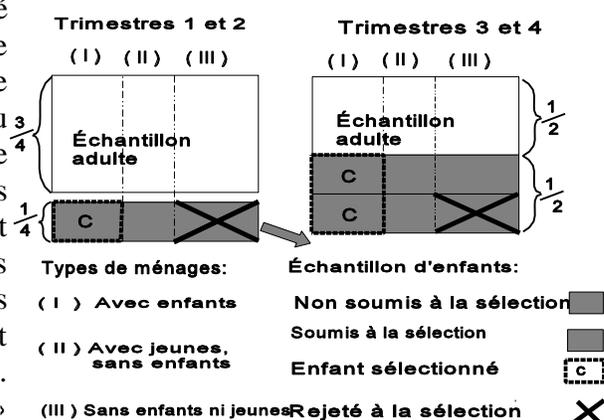
strates de l'EPA et dont les échantillons répétés sont habituellement des échantillons indépendants, répartis de manière identique et composés de quatre grappes chacun. Les quelques exceptions à cette règle ne devraient pas avoir d'incidence appréciable sur les résultats de l'enquête.

5.4 Intégration de l'ENSP et de l'Enquête longitudinale nationale sur les enfants (ELNE)

L'Enquête longitudinale nationale sur les enfants (ELNE) est une enquête-ménage qui suivra dans le temps un échantillon d'environ 25 000 enfants âgés de moins de 12 ans. L'ENSP et l'ELNE sont intégrées, ce qui veut dire que des données communes sur les enfants sont recueillies au moyen des deux enquêtes et que l'échantillon d'enfants de l'ENSP servira à produire les estimations de l'ELNE. Dans les provinces, on prévoit que l'ENSP fournira à l'ELNE un échantillon de 4 500 à 5 000 enfants. Pour obtenir cette taille d'échantillon, on fera remplir à tous les enfants de la famille (jusqu'à un maximum de quatre) le questionnaire détaillé des enfants, dans les ménages de l'ENSP où un enfant est choisi pour faire partie du panel.

À cause de contraintes liées au calendrier des activités, aucun enfant n'a été sélectionné pour le panel de l'ENSP avant la troisième période de collecte (ou trimestre). Cela a faussé la représentativité saisonnière des enfants dans le panel et a réduit la taille de l'échantillon correspondant. Afin d'accroître le rendement de l'échantillon des enfants, sans nuire à la représentativité saisonnière des autres membres du ménage durant les deux derniers trimestres, on a décidé de réattribuer à ces trimestres une partie de l'échantillon de l'ENSP des deux premiers trimestres. Étant donné que cette décision a été prise après les opérations d'échantillonnage décrites ci-haut, la réattribution a été appliquée aux ménages compris dans les grappes plutôt qu'aux grappes entières.

La figure 1 montre comment on a remanié la répartition de l'échantillon aux fins de l'intégration. Le carré de gauche représente une grappe attribuée au trimestre 1 ou 2, et celui de droite une grappe attribuée au trimestre 3 ou 4. Les ménages sont classés par type comme suit : (I) les ménages ayant des enfants; (II) les autres ménages comptant des jeunes (personnes âgées de moins de 25 ans); et (III) les ménages sans enfant ni jeune. L'échantillon comporte un volet «adultes» et un volet «enfants». Dans les ménages du volet «adultes», seules les personnes âgées de 12 ans et plus peuvent être



sélectionnées pour faire partie du panel. Dans les ménages du volet «enfants», seuls les enfants font partie du panel. S'il n'y a pas d'enfant, le ménage est rejeté à la sélection ou un membre (autre qu'un enfant) est choisi au hasard pour faire partie du panel.

Le quart de l'échantillon des trimestres 1 et 2 ainsi que la moitié de celui des trimestres 3 et 4 sont désignés comme des ménages «enfants». Les ménages «enfants» du trimestre 1 ou 2 seront en fait visités aux trimestres 3 et 4, respectivement. Sauf pour l'Île-du-Prince-Édouard, la méthode d'exclusion peut être appliquée strictement au volet «enfants» de l'échantillon. Lorsque le taux de sélection atteint 37,5 %, le tri s'applique à tous les ménages «enfants», ce qui n'est pas le cas lorsque le taux est plus faible. La figure 1 représente un taux de sélection de 25 %. Tous les ménages «enfants» des trimestres 1 et 2 ainsi que la moitié de ceux des trimestres 3 et 4 sont soumis au tri de sélection. Selon cette méthode, le panel comptera chaque trimestre à peu près le nombre de personnes qui ne sont pas des enfants. Cependant, il y aura des variations saisonnières du rendement de l'échantillon pour chaque type de ménage.

Pour des raisons d'ordre opérationnel, il n'y a aucune exclusion ni aucun déplacement des périodes de collecte dans les strates d'appartements, les strates de revenu élevé et les régions éloignées. À l'Île-du-Prince-Édouard, il n'y avait pas assez d'intervieweurs pour déplacer les périodes de collecte et il a donc fallu faire un tri de sélection à tous les trimestres. Dans ces cas, l'échantillon «enfants» est sélectionné uniquement à partir des grappes aux trimestres 3 et 4, ce qui cause une distorsion saisonnière de l'échantillon pour les personnes qui ne sont pas des enfants. De plus, à cause d'une erreur de programmation, aucun enfant âgé de 12 ans n'a été sélectionné pour faire partie du panel aux trimestres 1 et 2. Pour compenser, les probabilités de sélection des enfants âgés de 12 ans ont été ajustées aux trimestres 3 et 4; cependant, il en résulte que les enfants de 12 ans, tout comme ceux âgés de moins de 12 ans, ne sont pas représentés dans le panel aux deux premiers trimestres.

5.5 Plan d'échantillonnage au Québec

Au Québec, l'échantillon de l'ENSP a été prélevé parmi les logements qui ont participé à l'Enquête sociale et de santé (ESS), réalisée par Santé Québec en 1992-1993. Cette enquête a porté sur 16 010 logements qui ont été sélectionnés selon un plan d'échantillonnage à deux degrés semblable à celui de l'EPA. La province a été divisée géographiquement en recoupant quinze régions sanitaires avec quatre catégories de densité urbaine (la région métropolitaine de recensement de Montréal, les capitales régionales, les petites agglomérations urbaines et le secteur rural). Dans chaque région, les grappes ont été stratifiées selon des caractéristiques socio-économiques et sélectionnées à l'aide d'une méthode d'échantillonnage PPT. On a dénombré les grappes sélectionnées et on a prélevé des échantillons aléatoires de logements parmi celles-ci, à raison de 10 logements par grappe dans les grandes villes, et de 20 ou 30 logements ailleurs.

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

Santé Québec a fourni des renseignements non confidentiels qui ont permis de classer l'échantillon de l'ESS en quatre types de ménages : les ménages n'ayant qu'un seul membre; les ménages ayant des enfants; les autres ménages comptant des jeunes (personnes âgées de moins de 25 ans); et les autres (plus d'un membre, mais sans enfant ni jeune). Le type de ménages dans le cas des non répondants de l'ESS a été déterminé par le personnel de l'ENSP.

La taille de l'échantillon de l'ENSP a d'abord été répartie parmi les quatre catégories de densité urbaine. Afin d'éviter la concentration de l'échantillon à Montréal, la répartition a été proportionnelle à $\sqrt{(2W_h^2 + 1/4^2)}$, où W_h représente la part de la population pour la catégorie h , $h=1,2,3,4$. Dans chaque catégorie, on a tenté d'obtenir un sous-échantillon de l'ESS qui serait, en ce qui concerne le membre sélectionné pour faire partie du panel, proportionnel aux populations des quatre types de ménages. Pour ce faire, on a tiré de l'ESS un nombre suffisant de ménages pour donner le rendement voulu de ménages ayant des enfants (le groupe le plus sous-représenté), après quoi on a supprimé la part excédentaire des trois autres groupes de ménages. On a donc prélevé un échantillon initial supérieur de près de 50 % à l'échantillon requis. Après en avoir supprimé les deux tiers des ménages n'ayant qu'un seul membre, la moitié des autres ménages sans enfant ni jeune et le sixième des ménages ayant des jeunes mais pas d'enfant, l'objectif a été presque atteint.

Au Québec comme ailleurs, la représentation saisonnière, l'estimation de la variance et l'intégration avec l'ELNE ont eu une incidence sur le sous-échantillonnage. On a donc groupé des strates de l'ESS pour pouvoir former des échantillons répétés, les grappes de chacun couvrant les quatre trimestres (une grappe couvre deux trimestres dans les régions rurales et les petites agglomérations urbaines, car la taille des échantillons y est plus élevée). L'échantillon des ménages ayant des enfants a été réparti, selon un rapport de 3 à 2, entre un échantillon «adultes» et un échantillon «enfants», ces termes ayant la même signification que dans les autres provinces. Les ménages de l'échantillon «enfants» des trimestres 1 et 2 ont été réattribués aux trimestres 3 et 4. Pour les trimestres 3 et 4, l'échantillon des ménages sans enfant a lui aussi été réparti selon un rapport de 3 à 2, entre un échantillon «adultes» et un échantillon «enfants», étant donné que l'ENSP vise les occupants actuels des logements choisis pour l'ESS et que des changements seront survenus dans certains de ces logements.

6. Collecte des données

6.1 Conception du questionnaire et méthode de collecte des données

Les questions de l'ENSP ont été conçues pour être posées selon la méthode d'interview assistée par ordinateur (IAO), ce qui veut dire que l'ordre logique des questions a été programmé au fur et à mesure que celles-ci ont été élaborées. On a notamment spécifié

le genre de réponse voulue, les valeurs minimales et maximales, les vérifications en ligne liées aux questions et la suite à donner en cas de non-réponse.

L'IAO permet de contrôler l'interview d'après les déclarations du répondant. La possibilité de régler l'enchaînement des questions selon les réponses et de fixer des minimums et des maximums sont deux exemples importants de tels contrôles. Lorsqu'une réponse invalide est entrée, un message s'affiche à l'écran, ce qui permet au répondant et/ou à l'intervieweur d'obtenir un retour d'information immédiat pour corriger les incohérences. L'ajout automatique de la période de référence d'après la date du jour est un autre exemple d'amélioration. Il est possible de pré-insérer du texte ou des données selon les renseignements recueillis durant l'interview, ce qui permet à l'intervieweur de continuer son travail sans devoir revenir en arrière pour trouver des réponses antérieures. Cette possibilité permet notamment d'insérer le nom du répondant ou de faire les accords en genre à l'intérieur même des questions. Il est également possible de programmer des intervalles ou des réponses acceptables d'après les données recueillies durant l'interview. En d'autres mots, le questionnaire peut être adapté au répondant selon les données recueillies au moment de l'interview.

6.2 Essais

Un certain nombre d'essais ont été effectués avant la réalisation de l'enquête principale sur le terrain.

Des discussions dirigées en groupe ont eu lieu durant les étapes d'élaboration du questionnaire, afin de vérifier divers aspects de son contenu. Les principaux objectifs de ces discussions étaient de vérifier la clarté et la qualité des questions, les réactions des répondants aux sections jugées délicates (santé mentale, consommation d'alcool, etc.) et de déterminer la durée approximative des différentes sections.

Deux essais sur le terrain ont été effectués. Quatre des bureaux régionaux de Statistique Canada ont participé à ces essais, et les interviews ont été menées par des intervieweurs expérimentés de l'Enquête sur la population active. Les principaux objectifs de ces deux essais étaient encore une fois d'observer les réactions des répondants à l'enquête, d'estimer la durée des diverses sections du questionnaire et de déterminer les taux de réponse. La vérification des taux de réponse était particulièrement importante au Québec, puisque les répondants choisis avaient participé à l'enquête de Santé Québec quelques mois auparavant. Les opérations et les procédures sur le terrain, la formation des intervieweurs et le programme d'ordinateur (la version informatisée du questionnaire) ont aussi été soumis à des essais.

En plus des essais sur le terrain, le programme d'ordinateur a été soumis à des essais intensifs à l'interne pour le mettre au point et s'assurer que tous les cheminements

possibles étaient suivis correctement. La mise à l'essai du programme d'ordinateur s'est poursuivie de façon continue jusqu'au début de l'enquête principale.

6.3 Interviews

Les opérations de collecte des données ont été divisées en quatre trimestres (juin, août et novembre 1994 et mars 1995) et les interviews ont été menées par des intervieweurs de l'Enquête sur la population active (EPA) de Statistique Canada; ces intervieweurs sont des employés à temps partiel qui sont recrutés et formés expressément pour mener l'EPA selon la méthode d'interview assistée par ordinateur.

Le premier contact avec tous les répondants s'est fait au moyen d'une visite sur place, sauf pour un petit échantillon acheté par la province de la Colombie-Britannique, dans le cas duquel le premier contact s'est fait par téléphone selon la méthode de composition aléatoire (CA). Bon nombre des interviews débutées sur place ont été terminées par téléphone, soit parce que le répondant choisi n'était pas disponible au moment de la première visite, soit parce que l'interview était trop longue pour pouvoir l'achever en un seul contact. L'interview complète a duré en moyenne une heure auprès de chaque ménage.

Dans chaque logement, les renseignements sur tous les membres du ménage sont recueillis auprès d'un membre informé du ménage - il s'agit généralement de la personne qui se trouve à la maison au moment de la visite de l'intervieweur. On a recours à de telles déclarations par «personne interposée», qui représentent environ 55 % des renseignements recueillis, car des visites ou des appels répétés pour interviewer chaque répondant en personne coûteraient trop cher et exigeraient trop de temps.

Dans le cas des répondants choisis, les déclarations par personne interposée ont été permises seulement pour des raisons de maladie ou d'incapacité. Ces déclarations par personne interposée représentent 4 % des renseignements recueillis.

6.4 Surveillance et contrôle

Tous les intervieweurs de l'EPA travaillent sous la surveillance d'une équipe d'intervieweurs principaux qui sont chargés de veiller à ce que les intervieweurs connaissent bien les concepts et les méthodes de l'EPA et de ses nombreuses enquêtes supplémentaires, de contrôler périodiquement le travail des intervieweurs et de revoir les documents qu'ils remplissent. Les intervieweurs principaux travaillent à leur tour sous la surveillance des gestionnaires de programme de l'EPA, qui sont postés à chacun des huit bureaux régionaux de Statistique Canada.

Les intervieweurs principaux ont recontacté certains ménages après le troisième et le quatrième trimestres, afin de contrôler la qualité du travail des intervieweurs. Cette

opération consistait à vérifier la composition du ménage et à obtenir une évaluation du travail de l'intervieweur.

6.5 Non-réponse à l'ENSP

Les intervieweurs ont pour instructions de faire tous les efforts raisonnables pour interroger les membres des ménages admissibles à l'ENSP. Si une personne refuse dès le départ de participer à l'ENSP, le bureau régional envoie une lettre à l'adresse du logement pour souligner l'importance de l'enquête et de la coopération du ménage. Ensuite, l'intervieweur fait un deuxième appel (ou une deuxième visite). Lorsque le moment de l'appel (ou de la visite) de l'intervieweur est mal choisi, ce dernier fixe un rendez-vous pour rappeler ou revenir à un moment plus opportun. Lorsqu'il n'y a personne à la maison, de nombreuses autres tentatives sont faites pour contacter le ménage. En cas de non-réponse, les logements compris dans l'échantillon ne sont jamais remplacés par d'autres logements.

Chaque trimestre, après que toutes les tentatives ont été faites pour obtenir des interviews, il reste un petit nombre de ménages non répondants.

6.6 Suivi des cas de non-réponse

De nombreuses stratégies ont été adoptées pour réduire le nombre de cas de non-réponse. Avant le début de interviews, la taille de la tâche confiée à chaque intervieweur est limitée d'après les résultats des essais, afin de permettre un suivi efficace des cas de non-contact (c'est-à-dire pour éviter de surcharger les intervieweurs).

Dans le cadre des procédures d'interview, des moyens ont été proposés pour réduire le nombre de cas de non-contact, par exemple faire des visites à différentes heures de la journée en se rendant à d'autres logements ou sur le chemin de retour, parler à des voisins ou aux locataires pour connaître le nom des occupants d'un logement et obtenir un numéro de téléphone, etc.

Les cas de refus ont fait l'objet d'un suivi par les intervieweurs principaux, les surveillants de projet ou d'autres intervieweurs, afin de tenter de convaincre les répondants de participer à l'enquête.

Les questionnaires ont été produits dans les deux langues officielles et traduits en espagnol, en portugais, en chinois, en pendjabi et en italien pour tenter de réduire le nombre de cas de non-interview attribuables à des problèmes de langue.

Afin de maximiser le taux de réponse, un grand nombre de cas de non-réponse ont fait l'objet d'un suivi lors des trimestres subséquents.

7. Traitement des données

7.1 Saisie des données

Étant donné que l'ENSP a été menée selon la méthode d'IAO, la saisie faisait partie du processus de collecte des données. Les données recueillies durant l'interview étaient enregistrées directement dans un ordinateur portatif. Chaque question était affichée à l'écran de l'ordinateur. Une fois que la réponse à une question était entrée, la prochaine question apparaissait automatiquement à l'écran.

7.2 Vérification

Certaines étapes de vérification qui relèvent habituellement du bureau central ont été exécutées en direct par la voie du programme d'interview assistée par ordinateur (IAO) durant le processus de collecte des données. Les opérations de vérification pour déceler les valeurs aberrantes et les erreurs d'enchaînement ont été contrôlées à l'aide du système d'IAO. Il était possible de contrôler ce genre d'erreurs au moyen du système d'IAO, puisqu'on ne pouvait pas y entrer des valeurs incorrectes en tant que réponses, ni suivre un mauvais enchaînement des questions. Le système d'IAO, par exemple, faisait en sorte que toute question ne s'appliquant pas au répondant, et qui de ce fait ne devait pas avoir de réponse, restait sans réponse. Dans d'autres cas, des messages d'avertissement étaient donnés, mais aucune mesure corrective n'était prise lorsqu'un intervieweur entrait des réponses contradictoires entre les questions. Étant donné qu'aucune mesure corrective n'était prise en pareil cas, des vérifications ont été élaborées pour être exécutées au bureau central après la collecte des données. De façon général, les incohérences ont été corrigées en sélectionnant «non déclaré» pour l'une des variables en cause, ou les deux. Aucune imputation n'a été effectuée.

7.3 Codage

Dans le cas de plusieurs questions auxquelles il était possible de donner une réponse écrite, les renseignements écrits étaient codés en de nouvelles catégories distinctes ou en une catégorie inscrite si ces renseignements faisaient double emploi avec une catégorie existante. Dans les cas où cela était possible (p. ex. la profession, l'industrie, les maladies), le codage correspondait aux systèmes de classification type utilisés dans le Recensement de la population ou dans d'autres enquêtes de Statistique Canada, par exemple l'Enquête sur la santé et les limitations d'activités et le cycle 6 de l'Enquête sociale générale.

7.4 Création de variables calculées

Un certain nombre de variables du fichier de microdonnées ont été calculées en utilisant des postes du questionnaire de l'ENSP pour faciliter l'analyse des données. Le nom des

variables dérivées est habituellement précédé des lettres DV et suivi de caractères se rapportant au numéro de la question ou au sujet. Dans certains cas, les variables dérivées sont simples et elles ont été obtenues en groupant des catégories. Dans d'autres cas, plusieurs variables ont été combinées pour former une nouvelle variable. Les détails du calcul de ces variables figurent à l'annexe D.

7.5 Facteurs de pondération

Le principe sur lequel repose l'estimation pour un échantillon probabiliste comme celui de l'ENSP veut que chacune des personnes faisant partie de l'échantillon «représente», à part elle-même, plusieurs autres personnes qui n'en font pas partie. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne représente 50 membres de la population. Dans la terminologie de ce texte, on peut dire que chaque personne a un coefficient de pondération (ou poids) de 50.

La phase de pondération est l'étape où l'on calcule ce poids pour chaque personne. Ce poids est enregistré dans le fichier de microdonnées et il faut l'utiliser pour produire des estimations valables à partir des résultats de l'enquête. Par exemple, si l'on veut estimer le nombre de personnes au Canada qui fument tous les jours (voir la question SMOK-Q2 à la section 9.2), on choisit les enregistrements se rapportant aux personnes qui possèdent cette caractéristique et on fait la somme des poids inscrits sur ces enregistrements.

La méthode employée pour calculer ces poids est décrite en détail au chapitre 11.

7.6 Suppression de renseignements confidentiels

Il convient de mentionner que le fichier de microdonnées «à grande diffusion» décrit ci-dessus diffère à bien des égards du fichier «principal» de l'enquête conservé par Statistique Canada. Ces écarts sont dus à des mesures prises pour protéger l'anonymat des répondants à l'enquête. Les utilisateurs qui désirent obtenir des renseignements exclus du fichier de microdonnées peuvent acheter des totalisations personnalisées. Les estimations produites seront communiquées à l'utilisateur, à la condition qu'elles soient conformes aux lignes directrices pour l'analyse et la diffusion décrites au chapitre 9 de ce guide.

8. Qualité des données

8.1 Taux de réponse

Deux éléments propres à l'ENSP ont compliqué le calcul des taux de réponse. Le premier problème est survenu à la suite de l'utilisation de la méthode d'exclusion. Rappelons qu'un certain pourcentage de logements avaient été identifiés comme étant éligibles au rejet (voir la section 5.2 pour plus de détails). Les ménages qui ont finalement été rejetés ne sont pas comptés dans les estimations d'enquête. Ils sont cependant considérés comme des ménages répondants étant donné qu'ils ont fourni l'information requise par l'ENSP. Les ménages éligibles au rejet pour lesquels on n'a pas obtenu de réponse sont considérés comme des ménages non-répondants (tout comme les ménages non éligibles au rejet pour lesquels on n'a pas obtenu de réponse).

Deuxièmement, le taux de réponse au niveau des personnes sélectionnées a été compliqué par l'intégration de l'ENSP avec l'ENLE. Rappelons que dans certains logements pré-identifiés, la présence d'un enfant de moins de douze ans faisait en sorte que la personne sélectionnée était un enfant et que celui-ci recevait le questionnaire de l'ENLE. Dans ces cas, personne ne répondait au questionnaire de la personne sélectionnée de l'ENSP. Pour cette raison, ces logements sont classés hors du champ de l'enquête aux fins de calcul du taux de réponse de l'ENSP au niveau des personnes sélectionnées.

Ce qui suit décrit comment les taux de réponse au niveau des ménages et au niveau des personnes sélectionnées ont été calculés. Il convient de noter que les logements hors du champ de l'enquête (les logements vacants ou abandonnés, les logements en construction, ainsi que les ménages n'étant pas éligibles à faire partie de l'échantillon) n'ont été utilisés dans aucuns calculs.

Taux de réponse au niveau des ménages

taux de réponse des ménages = $\frac{\text{nombre de ménages répondants, incluant les ménages rejetés}}{\text{tous les ménages faisant partie du champ de l'enquête}}$

Un ménage répondant non-rejeté était un ménage où, *au minimum*, un membre avait répondu à la composante générale du questionnaire. Le taux de réponse de l'ENSP au niveau des ménages pour l'ensemble du Canada était de **88.7%**. Au niveau provincial, ce taux variait de 85.2% en Ontario à 93.2% en Alberta.

Taux de réponse au niveau des personnes sélectionnées

Le taux de réponse au niveau des personnes sélectionnées peut être vu comme le rapport

GUIDE DU FICHIER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

entre le nombre de questionnaires sur la santé qui ont été *effectivement* remplis et le nombre de questionnaires sur la santé qui *auraient dû* être remplis.

$$\text{taux de rép. des pers. sél.} = \frac{\text{nombre de questionnaires sur la santé remplis}}{\text{nombre de personnes éligibles à répondre au questionnaire sur la santé}}$$

où le nombre de personnes éligibles à répondre au questionnaire sur la santé est égal au nombre de ménages répondants non-rejetés duquel on retranche le nombre de logements où un enfant de moins de douze ans était la personne sélectionnée.

Le taux de réponse au niveau des personnes sélectionnées était de **96.1%** pour l'ensemble du Canada, et variait de 94.7% en Nouvelle-Écosse à 97.6% en Saskatchewan.

Il convient de souligner qu'à cause des complications décrites précédemment, la multiplication des deux taux donne un résultat dénué de sens. L'information utilisée pour calculer chacun de ces taux est différente, ce qui fait qu'un taux combiné ne peut être déterminé.

Information pertinente au calcul du taux réponse:

Nombre de répondants au niveau du ménage:	20725
Nombre de répondants - personne choisie:	17626
Nombre de ménages exclus:	3447
Nombre de logements - enfant choisi:	2383
Nombre de non-réponse au niveau du ménage:	3091
Nombre de non-réponse - personne choisie:	716
Nombre de ménages:	4512

Calcul du taux de réponse pour le ménage;

$$\text{Taux MM} = \frac{20725 + 3447}{20725 + 3447 + 3091} = \frac{24172}{27263} = 88.7\%$$

Calcul du taux de réponse pour les personnes choisies:

$$\text{Taux PC} = \frac{17626}{20725 - 2383} = \frac{17626}{18342} = 96.1\%$$

8.2 Erreurs relatives à l'enquête

L'enquête permet de produire des estimations fondées sur les données portant sur un échantillon et recueillies auprès de ses membres. On aurait pu obtenir des estimations quelque peu différentes si on avait procédé à un recensement complet en faisant appel au même questionnaire et aux mêmes intervieweurs, surveillants, méthodes de traitement, etc., que pour cette enquête. L'écart entre les estimations découlant de l'échantillon et celles tirées d'un dénombrement complet effectué dans des conditions semblables est appelé erreur d'échantillonnage des estimations.

Des erreurs qui ne se rapportent pas à l'échantillonnage peuvent se produire à presque toutes les étapes d'une enquête. Il se peut que les intervieweurs comprennent mal les instructions, que les répondants fassent des erreurs en répondant aux questions, que des réponses soient mal entrées dans l'ordinateur et que des erreurs se produisent au moment du traitement et de la totalisation des données. Ces erreurs sont toutes des exemples d'erreurs non dues à l'échantillonnage.

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations tirées des données de l'enquête. Toutefois, les erreurs qui se produisent systématiquement contribueront à des biais dans les estimations de l'enquête. On a consacré beaucoup de temps et d'efforts à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été appliquées à chaque étape du cycle de collecte et de traitement des données afin de contrôler la qualité des données. On a notamment fait appel à des intervieweurs hautement qualifiés, donné à ceux-ci une formation poussée sur les méthodes d'enquête et le questionnaire, observé les intervieweurs afin de déceler les problèmes liés à la version IAO du questionnaire ou à la compréhension des instructions, et adopté des procédures pour veiller à réduire au minimum les erreurs de collecte de données.

L'effet de la non-réponse sur les résultats de l'enquête constitue une source importante d'erreurs non dues à l'échantillonnage dans les enquêtes. L'ampleur de la non-réponse varie d'une non-réponse partielle (le fait de ne pas répondre à une ou plusieurs questions) à une non-réponse totale. Dans le cas de l'ENSP, il n'y a presque pas eu de non-réponse partielle; une fois le questionnaire débuté, les répondants avaient tendance à le terminer. Il y a eu non-réponse totale lorsque l'intervieweur a été incapable de contacter le répondant, qu'aucun membre du ménage n'a pu fournir les renseignements voulus ou que le répondant a refusé de participer à l'enquête. On a traité les cas de non-réponse totale en ajustant le poids des ménages qui ont répondu à l'enquête afin de compenser pour ceux qui n'ont pas répondu.

Dans la plupart des cas, il y a eu non-réponse partielle à l'enquête lorsque le répondant n'a pas compris ou a mal interprété une question, qu'il a refusé de répondre à une question,

GUIDE DU FICHIER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

qu'il n'a pas pu se souvenir du renseignement demandé ou qu'il n'a pas pu fournir de renseignement sans personne interposée.

Étant donné que les estimations d'une enquête par sondage comportent inévitablement une erreur d'échantillonnage, de bonnes méthodes statistiques exigent que les chercheurs fournissent aux utilisateurs une certaine indication de la grandeur de cette erreur d'échantillonnage. Cette partie du guide décrit les mesures de l'erreur d'échantillonnage que Statistique Canada utilise couramment et dont il conseille vivement l'emploi aux utilisateurs qui produisent des estimations à partir de ce fichier de microdonnées.

La mesure de l'importance éventuelle des erreurs d'échantillonnage est fondée sur l'erreur-type des estimations tirées des résultats de l'enquête. Cependant, en raison de la diversité des estimations que l'on peut tirer 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 (c.v.) 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 qu'à partir des résultats de l'enquête, on estime que 24 % des Canadiens âgés de 12 ans et plus fument des cigarettes tous les jours et qu'on établit à .003 l'erreur-type de cette estimation. Le coefficient de variation serait alors calculé comme suit :

$$\left(\frac{.003}{.24} \right) \times 100\% = 1.14\%$$

9. Lignes directrices pour la totalisation, l'analyse et la diffusion

Cette section du guide décrit les lignes directrices que doivent suivre les utilisateurs qui totalisent, analysent, publient ou diffusent de quelque façon que ce soit des données tirées des bandes de microdonnées de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs d'arriver aux mêmes chiffres que Statistique Canada, tout en étant capables d'obtenir, conformément à ces lignes directrices, des chiffres qui n'ont pas été publiés.

9.1 Lignes directrices pour l'arrondissement

Afin que les estimations destinées à la publication ou à toute autre forme de diffusion qui sont tirées de ces bandes de microdonnées correspondent à celles de Statistique Canada, nous conseillons vivement aux utilisateurs de se conformer aux lignes directrices suivantes pour l'arrondissement de ces estimations :

- a) Les estimations dans le corps d'un tableau statistique doivent être arrondies à la centaine près au moyen de la technique d'arrondissement classique. 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, on augmente de 1 la valeur du dernier chiffre à conserver. Par exemple, lorsqu'on veut arrondir à la centaine près de la façon classique, si les deux derniers chiffres se situent entre 00 et 49, il faut les remplacer par 00 et laisser le chiffre précédent (le chiffre des centaines) tel quel. Si les deux derniers chiffres se situent entre 50 et 99, on les remplace par 00 et on augmente de 1 le chiffre précédent.
- b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques doivent être calculés à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement classique.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir d'éléments non arrondis (c'est-à-dire les numérateurs et/ou les dénominateurs), puis arrondis à une décimale au moyen de la technique d'arrondissement classique. Lorsqu'on veut arrondir à un seul chiffre par cette technique, 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, on augmente de 1 le dernier chiffre à conserver.
- d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près (ou à la décimale près) en employant la technique d'arrondissement classique.

- e) Si, en raison de limitations d'ordre technique ou de toute autre nature, on utilise une autre technique que l'arrondissement classique, de sorte que les estimations à publier ou à diffuser sous une forme quelconque diffèrent des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.
- f) Des estimations non arrondies ne doivent en aucun cas être publiées ou diffusées de quelque façon que ce soit par les utilisateurs. Des estimations qui ne sont pas arrondies laissent supposer qu'elles sont beaucoup plus précises qu'elles ne 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'ENSP n'était pas autopondéré, c'est-à-dire que les personnes faisant partie l'échantillon n'ont pas toutes le même poids d'échantillonnage. Pour produire des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent employer le poids d'échantillonnage approprié.

Sinon, les estimations calculées à partir des bandes de microdonnées ne pourront être considérées comme représentatives de la population observée et ne correspondront pas à celles de Statistique Canada.

Les utilisateurs doivent également noter qu'en raison de la manière dont ils traitent le champ du poids, certains logiciels ne permettent peut-être pas d'obtenir des estimations qui correspondent exactement à celles de Statistique Canada.

9.2.1 Définitions des genres d'estimations : de type nominal par opposition à quantitatives

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'ENSP, il est utile de décrire les deux principaux genres d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites à partir du fichier de microdonnées de l'Enquête nationale sur la santé de la population.

Estimations de type nominal

Les estimations de type nominal sont des estimations du nombre ou du pourcentage de personnes dans la population visée par l'enquête qui possèdent certaines caractéristiques ou qui font partie d'une catégorie définie. Le nombre de personnes qui fument tous les jours est un exemple d'estimation de ce genre. On peut aussi utiliser l'expression «estimation d'un agrégat» pour parler d'une estimation du nombre de personnes qui possèdent une caractéristique donnée.

Exemple de question de type nominal :

SMOK-Q2 : Actuellement, est-ce que ... fume(z) des cigarettes tous les jours, à l'occasion ou jamais?

- Tous les jours
- À l'occasion
- Jamais

Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures de tendance centrale de quantités basées sur tous les membres de la population observée ou sur certains d'entre eux. Elles comprennent aussi explicitement des estimations de la forme \hat{X} / \hat{Y} , où \hat{X} est une estimation de la quantité totale pour la population observée et \hat{Y} une estimation du nombre de personnes dans la population observée qui contribuent à cette quantité totale.

Un exemple d'estimation quantitative est le nombre moyen de cigarettes que fument par jour les personnes qui fument tous les jours. Le numérateur correspond à une estimation du nombre total de cigarettes que fument par jour les personnes qui fument tous les jours, et le dénominateur à une estimation du nombre de personnes qui fument tous les jours.

Exemple de question quantitative :

SMOK-Q4 : Actuellement, combien de cigarettes est-ce que ... fume(z) chaque jour?

|_|_| Nombre de cigarettes

9.2.2 Totalisation d'estimations de type nominal

On peut obtenir des estimations du nombre de personnes qui possèdent une caractéristique donnée à partir du fichier de microdonnées en additionnant les poids finals de tous les enregistrements contenant la ou les caractéristiques voulues. On obtient les proportions et les rapports de la forme \hat{X} / \hat{Y} comme suit:

- a) en additionnant les poids finals des enregistrements contenant la caractéristique voulue pour le numérateur (\hat{X}),
- b) en additionnant les poids finals des enregistrements contenant la

- caractéristique voulue pour le dénominateur (\hat{Y}),
- c) en divisant l'estimation du numérateur par celle du dénominateur.

9.2.3 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable voulue par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements qui nous intéressent. Par exemple, pour obtenir une estimation du nombre total de cigarettes que fument chaque jour les personnes qui fument tous les jours, on multiplie la valeur déclarée à la question SMOK-Q4 par le poids final de l'enregistrement, puis on additionne ce résultat pour tous les enregistrements dont la réponse à la question SMOK-Q2 est «tous les jours».

Afin d'obtenir une moyenne pondérée exprimée sous la forme \hat{X} / \hat{Y} , le numérateur (\hat{X}) est calculé comme une estimation quantitative et le dénominateur (\hat{Y}) comme une estimation de type nominal. Par exemple, pour estimer le nombre moyen de cigarettes que fument par jour les personnes qui fument tous les jours, il faut :

- a) estimer le nombre total de cigarettes que fument par jour les personnes qui fument tous les jours, selon la méthode décrite ci-haut,
- b) estimer le nombre de personnes qui fument tous les jours en additionnant les poids finals de tous les enregistrements dont la réponse à la question SMOK-Q2 est «tous les jours», puis
- c) diviser l'estimation obtenue en «a» par celle calculée en «b».

9.3 Lignes directrices pour l'analyse statistique

L'Enquête nationale sur la santé de la population est basée sur un plan complexe qui comporte une stratification et de multiples degrés de sélection, ainsi que des probabilités inégales de sélection des répondants. L'utilisation de données provenant d'enquêtes aussi complexes pose des difficultés aux analystes, car le plan d'enquête et les probabilités de sélection influent sur les méthodes d'estimation et de calcul de la variance à utiliser.

Bien que de nombreuses méthodes d'analyse intégrées aux progiciels statistiques permettent d'utiliser des poids, le sens ou la définition du poids diffère de celui qui convient à une enquête par sondage. Par conséquent, si les estimations faites au moyen de ces progiciels sont exactes dans bien des cas, les variances calculées n'ont pratiquement aucune signification.

Dans le cas de nombreuses techniques d'analyse (par exemple la régression linéaire, la régression logistique, l'analyse de variance), il existe un moyen de rendre l'application des logiciels courants plus significative. Si l'on transforme les poids des enregistrements contenus dans le fichier de manière à ce que le poids moyen soit un (1), les résultats produits par les logiciels courants seront plus raisonnables et tiendront compte des probabilités inégales de sélection, bien qu'ils continueront à ne faire aucun cas de la stratification et de la répartition en grappes du plan d'échantillonnage. On peut effectuer cette transformation en employant dans l'analyse un poids égal au poids original divisé par la moyenne des poids originaux des unités d'échantillonnage (personnes) qui contribuent à l'estimateur en question.

Afin de permettre d'évaluer la qualité des estimations totalisées, Statistique Canada a produit un ensemble de Tables de variabilité d'échantillonnage approximative (couramment appelées «tables de c.v.») pour l'ENSP. On peut employer ces tables afin d'obtenir des coefficients de variation approximatifs pour les estimations et les proportions de type nominal. Pour plus de détails, voir le chapitre 10.

9.4 Lignes directrices pour la diffusion

Avant de diffuser et/ou de publier des estimations tirées de ces bandes de microdonnées, les utilisateurs doivent d'abord déterminer le nombre de répondants qui ont fourni les données entrant dans le calcul de l'estimation. Si ce nombre est inférieur à 30, l'estimation pondérée ne doit pas être diffusée, quelle que soit la valeur de son coefficient de variation. Pour les estimations pondérées basées sur des échantillons composés de 30 répondants ou plus, les utilisateurs doivent calculer le coefficient de variation de l'estimation arrondi et suivre les lignes directrices ci-dessous.

Lignes directrices relatives à la variabilité d'échantillonnage

Type d'estimation	c.v. (en %)	Lignes directrices
1. Sans réserve	0.0 à 16.5	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
2. Sous réserve	16.6 à 25.0	On peut envisager une diffusion générale non restreinte des estimations, en y joignant une mise en garde aux utilisateurs suivants quant à la variabilité d'échantillonnage élevée liée aux estimations. Les estimations de ce genre doivent être identifiées par la lettre Q (ou d'une autre manière semblable).
3. Confidentielle	25.1 à 33.3	On ne peut envisager une diffusion générale non restreinte des estimations que si la variabilité d'échantillonnage est obtenue au moyen d'une méthode de calcul exact de la variance. À défaut de variances exactes, il faut supprimer ces estimations des tableaux statistiques et les remplacer par des tirets (---).
4. Diffusion interdite	33.4 ou plus	Les estimations ne peuvent être diffusées sous aucune forme NI dans aucun cas. Dans les tableaux statistiques, ces estimations doivent être supprimées et remplacées par des tirets (---).

10. Tables de variabilité d'échantillonnage approximative

Afin de permettre aux utilisateurs d'avoir facilement accès à des coefficients de variation qui s'appliqueraient à une multitude d'estimations de type nominal obtenues à partir de ce fichier de microdonnées, Statistique Canada a produit un ensemble de Tables de variabilité d'échantillonnage approximative. Ces «tables à consulter» permettent aux utilisateurs d'obtenir un coefficient de variation approximatif selon la taille de l'estimation calculée à partir des données de l'enquête.

Les coefficients de variation (c.v.) dans ces tables sont calculés en employant la formule de la variance utilisée pour l'échantillonnage aléatoire simple et en y incorporant un facteur qui reflète la structure en grappes à plusieurs degrés du plan d'échantillonnage. Pour obtenir ce facteur, appelé effet du plan, on a d'abord calculé les effets du plan pour une vaste gamme de caractéristiques, puis on a choisi parmi les nombres ainsi obtenus une valeur modérée à employer dans les tables à consulter qui s'appliquerait ensuite à l'ensemble complet des caractéristiques.

Les quatre tableaux suivants montrent les effets du plan, la taille des échantillons et les chiffres de population, selon la province, qui ont servi à produire les Tables de variabilité d'échantillonnage approximative.

**Données d'entrée des tables de variabilité d'échantillonnage,
Canada et provinces, membres du ménage, tous les âges**

PROVINCE	EFFET DU PLAN	TAILLE DE L'ÉCHAN-TILLON	POPULATION
Terre-Neuve	1.39	3511	573863
Île-du-Prince-Édouard	1.28	3106	132727
Nouvelle-Écosse	1.36	3071	909257
Nouveau-Brunswick	1.34	3607	742680
Québec	1.69	8461	7124848
Ontario	1.67	17221	10824871
Manitoba	2.32	4744	1074307
Saskatchewan	1.33	3161	969226
Alberta	1.41	4487	2657513
Colombie-Britannique	1.83	7070	3608380
Canada	2.12	58439	28617677

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

Données d'entrée des tables de variabilité d'échantillonnage pour le Canada et les provinces, personnes choisies, 12 ans et plus

PROVINCE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION
Terre-Neuve	1.04	918	483363
Île-du-Prince-Édouard	0.97	899	109603
Nouvelle-Écosse	1.01	911	763944
Nouveau-Brunswick	1.07	1111	626303
Québec	1.22	2581	6029670
Ontario	1.37	5187	9050016
Manitoba	1.61	1420	890750
Saskatchewan	1.03	1005	792049
Alberta	1.05	1310	2166102
Colombie-Britannique	1.54	2284	3036798
Canada	1.64	17626	23948603

Données d'entrée des tables de variabilité de l'échantillonnage selon le groupe d'âge, membres du ménage, tous les âges

GROUPE D'ÂGE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION
0-11	2.03	11477	4669074
12-24	2.13	11706	5111949
25-44	1.79	18922	9619826
45-64	1.84	11032	5965860
65+	1.96	5302	3250967

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

Données d'entrée des tables de variabilité de l'échantillonnage selon le groupe d'âge, personnes choisies, 12 ans et plus

GROUPE D'ÂGE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION
12-24	1.86	3242	5111949
25-44	1.60	6790	9619826
45-64	1.50	4451	5965860
65+	1.28	3143	3250967

Tous les coefficients de variation sont approximatifs dans les Tables de variabilité d'échantillonnage approximative et ils ne doivent donc pas être considérés comme des valeurs officielles. Il est possible d'obtenir de Statistique Canada, contre recouvrement des frais, des estimations de la variance réelle de variables particulières. L'utilisation d'estimations de la variance réelle peut permettre aux utilisateurs de diffuser des estimations qui ne pourraient être diffusées autrement, c'est-à-dire des estimations dont les coefficients de variation se situent dans la catégorie «confidentielle»

Rappel : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée ne doit pas être diffusée, quelle que soit la valeur de son coefficient de variation, car les formules employées pour estimer la variance ne sont pas valables pour des échantillons de moindre taille.

10.1 Comment utiliser les tables de c.v. pour les estimations de type nominal

Les règles suivantes devraient permettre à l'utilisateur de calculer à partir des Tables de variabilité d'échantillonnage les coefficients de variation approximatifs d'estimations relatives au nombre, à la proportion ou au pourcentage de personnes dans la population observée qui possèdent une caractéristique donnée ainsi que des rapports et des écarts entre ces estimations.

Règle 1 : Estimations du nombre de personnes 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 correspondant à la région géographique appropriée, il faut repérer le nombre estimé dans la colonne d'extrême gauche (intitulée «Numérateur du pourcentage») et

suivre les astérisques (s'il y en a) de gauche à droite jusqu'au premier nombre. Ce nombre constitue le coefficient de variation approximatif.

Règle 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Le coefficient de variation d'une proportion ou d'un pourcentage estimé dépend à la fois de la grandeur de cette proportion ou de ce pourcentage et de la grandeur du total sur lequel est fondé cette proportion ou ce pourcentage. Les proportions ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque cette proportion ou ce pourcentage est fondé sur un sous-ensemble de la population. Cela est dû au fait que les coefficients de variation des estimations du dernier type sont basés sur le chiffre le plus élevé dans une rangée d'une table particulière, tandis que les coefficients de variation des estimations du premier type sont basés sur un chiffre quelconque de cette même rangée (pas nécessairement le plus élevé). (Il convient de noter que dans les tables, la valeur des coefficients de variation décroît de gauche à droite sur une même ligne.) Par exemple, la proportion estimative de personnes qui fument tous les jours est plus fiable que le nombre estimatif de personnes qui fument tous les jours.

Lorsque la proportion ou le pourcentage est fondé sur la population totale de la région géographique à laquelle la table s'applique, le coefficient de variation de la proportion ou du pourcentage est égal à celui du numérateur de la proportion ou du pourcentage. Dans ce cas-ci, on peut appliquer la règle 1.

Lorsque la proportion ou le pourcentage est fondé sur un sous-ensemble de la population totale (p. ex. les personnes qui fument), il faut se reporter à la proportion ou au pourcentage (haut de la table) et au numérateur de la proportion ou du pourcentage (côté gauche de la table). Le coefficient de variation se trouve à l'intersection de la ligne et de la colonne appropriée.

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

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

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

où \hat{X}_1 représente l'estimation 1, \hat{X}_2 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 donne un résultat exact pour ce qui est de la différence entre des caractéristiques distinctes et non corrélées, mais elle ne donne que des résultats approximatifs dans les autres cas.

Règle 4 : Estimations de rapports

Si le numérateur est un sous-ensemble du dénominateur, il faut convertir le rapport en pourcentage et appliquer la règle 2. Ce serait le cas, par exemple, si le dénominateur est le nombre de personnes qui fument et le numérateur est le nombre de personnes qui fument tous les jours parmi celles qui fument.

Si le numérateur n'est pas un sous-ensemble du dénominateur (par exemple, le rapport du nombre de personnes qui fument tous les jours ou à l'occasion au nombre de personnes qui ne fument pas du tout), l'écart-type du rapport entre les estimations est à peu près égal à la racine carrée de la somme des carrés de chaque coefficient de variation pris séparément multipliée par \hat{R} , où \hat{R} est le rapport des estimations ($\hat{R} = \hat{X}_1 / \hat{X}_2$). L'erreur-type d'un rapport est donc :

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

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

Le coefficient de variation de \hat{R} est donné par $\sigma_{\hat{R}} / \hat{R} = \sqrt{\alpha_1^2 + \alpha_2^2}$. La formule tend à surestimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés positivement et à sous-estimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés négativement.

Règle 5 : Estimations des différences entre des rapports

On commence par calculer les coefficients de variation des deux rapports au moyen de la règle 4, puis le coefficient de variation de leur différence au moyen de la règle 3.

10.2 Exemples d'utilisation des tables de c.v. pour des estimations de type nominal

Les exemples «réels» suivants ont pour but d'aider les utilisateurs à appliquer les règles décrites ci-haut.

Exemple 1 : Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Supposons qu'un utilisateur estime à 5 958 122 le nombre de personnes qui fument tous les jours au Canada. Comment l'utilisateur fait-il pour déterminer le coefficient de variation de cette estimation?

- 1) Se reporter à la table de c.v. pour le CANADA pour les personnes choisies.
- 2) L'agrégat estimé (5 958 122) ne figure pas dans la colonne de gauche (la colonne «Numérateur du pourcentage»); il faut donc utiliser le nombre qui s'en rapproche le plus, soit 6 000 000.
- 3) Le coefficient de variation d'un agrégat estimé (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 1,6 %.
- 4) Le coefficient de variation approximatif de l'estimation est donc 1,6 %. Par conséquent, l'estimation selon laquelle 5 958 122 personnes fument tous les jours peut être diffusée sans réserve.

Exemple 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Supposons qu'un utilisateur estime à $5\,958\,122/8\,937\,183 = 66,7\%$ le pourcentage de personnes qui fument tous les jours au Canada. Comment l'utilisateur fait-il pour déterminer le coefficient de variation de cette estimation?

- 1) Se reporter à la table de c.v. pour le CANADA pour les personnes choisies.
- 2) Parce que l'estimation est un pourcentage basé sur un sous-ensemble de la population totale (c.-à-d. les personnes qui fument tous les jours ou à l'occasion), il faut utiliser à la fois le pourcentage (66,7 %) et la partie numérateur du pourcentage (5 958 122) pour déterminer le coefficient de variation.
- 3) Le numérateur (5 958 122) ne figure pas dans la colonne de gauche (la colonne «Numérateur du pourcentage»); il faut donc utiliser le nombre qui s'en rapproche le plus, soit 6 000 000. De même, l'estimation du pourcentage ne figure pas parmi les

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

en-têtes de colonnes; il faut donc utiliser le nombre qui s'en rapproche le plus, soit 70,0 %.

- 4) Le nombre qui se trouve à l'intersection de la ligne et de la colonne utilisées, soit 1,1 %, est le coefficient de variation (exprimé en pourcentage) à employer.
- 5) Le coefficient de variation de l'estimation est donc 1,1 %. Par conséquent, l'estimation selon laquelle 66,7 % des gens qui fument le font tous les jours peut être diffusée sans réserve.

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

Supposons qu'un utilisateur estime que, parmi les personnes qui fument tous les jours, 2 859 899/5 958 122 = 48 % d'entre elles fument 10 cigarettes ou plus par jour (estimation 1) et que, parmi les personnes qui fument à l'occasion ou celles qui ne fument pas du tout mais qui ont déjà fumé tous les jours, 3 160 514/4 329 471 = 73 % d'entre elles fument ou ont déjà fumé 10 cigarettes ou plus par jour (estimation 2). Il est à noter que ces estimations sont basées sur les réponses aux questions SMOK-Q2, SMOK-Q4, SMOK-Q4A, SMOK-Q5 et SMOK-Q7. Comment l'utilisateur fait-il pour déterminer le coefficient de variation de la différence entre ces deux estimations?

- 1) À l'aide de la table de c.v. pour le CANADA, personnes choisies utilisée de la même façon que dans l'exemple 2, vous établissez à 1,9 % le c.v. de l'estimation 1 (exprimé en pourcentage) et à 1,5 % le c.v. de l'estimation 2 (exprimé en pourcentage).
- 2) Selon la règle 3, l'erreur-type pour une différence ($\hat{d} = \hat{X}_2 - \hat{X}_1$) 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.

L'erreur-type de la différence $\hat{d} = (.73 - .48) = .25$ est donc :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(.48)(.019)]^2 + [(.73)(.015)]^2} \\ &= .014\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = .014 / .25 = 0.056$.

- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 5,6 % (exprimé en pourcentage). Par conséquent, cette estimation peut être publiée sans réserve.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 5 958 122 le nombre de personnes qui fument tous les jours et à 1 732 412 le nombre de celles qui fument à l'occasion. L'utilisateur veut comparer ces deux estimations sous la forme d'un rapport. Comment fait-il pour déterminer le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation de rapport, où le numérateur de l'estimation ($= \hat{X}_1$) est le nombre de personnes qui fument à l'occasion. Le dénominateur de l'estimation ($= \hat{X}_2$) est le nombre de personnes qui fument tous les jours.
- 2) Se reporter à la table de c.v. pour le CANADA pour les personnes choisies.
- 3) Le numérateur de cette estimation de rapport est 1 732 412. Le chiffre qui se rapproche le plus de ce nombre est 1 500 000. Le coefficient de variation de cette estimation (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 3,7 %.
- 4) Le dénominateur de cette estimation de rapport est 5 958 122. Le chiffre qui se rapproche le plus de ce nombre est 6 000 000. Le coefficient de variation de cette estimation (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 1,6 %.
- 5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, soit

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

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

C'est-à-dire,

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(.037)^2 + (.016)^2} \\ &= 0.040\end{aligned}$$

Le rapport des personnes qui fument occasionnellement à celles qui fument tous les jours est 1 732 412/5 958 122, soit 0.29:1. Le coefficient de variation de cette estimation est 4,0 % (exprimé en pourcentage); l'estimation peut donc être diffusée sans réserve.

10.3 Comment utiliser les tables de c.v. pour calculer les limites de confiance

Bien que les coefficients de variation soient largement utilisés, l'intervalle de confiance d'une estimation donne une mesure intuitive plus significative de l'erreur d'échantillonnage. Un intervalle de confiance est une façon d'énoncer la probabilité que la valeur vraie de la population se situe dans une plage de valeurs données. Par exemple, un intervalle de confiance de 95 % peut être décrit comme suit :

Si l'échantillonnage de la population se répète à l'infini, chacun des échantillons donnant un nouvel intervalle de confiance pour une estimation, l'intervalle contiendra la valeur vraie de la population dans 95% des cas.

Une fois déterminée l'erreur-type d'une estimation, on peut calculer des intervalles de confiance pour les estimations en partant de l'hypothèse qu'en procédant à un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique de la population sont réparties selon une distribution normale 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 de 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. On appelle ces différents degrés de confiance des niveaux de confiance.

Les intervalles de confiance d'une estimation, \hat{X} , sont généralement exprimés sous forme de deux nombres, l'un étant inférieur à l'estimation et l'autre supérieur à celle-ci, sous la forme $(\hat{X}-k, \hat{X}+k)$, où k varie selon le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

On peut calculer directement les intervalles de confiance d'une estimation à partir des Tables de variabilité d'échantillonnage approximative, en trouvant d'abord dans la table appropriée le coefficient de variation de l'estimation \hat{X} , puis en utilisant la formule suivante pour obtenir l'intervalle de confiance CI correspondant :

$$CI_X = [\hat{X} - t\hat{X}\alpha_{\hat{X}}, \hat{X} + t\hat{X}\alpha_{\hat{X}}]$$

où $\alpha_{\hat{X}}$ est le coefficient de variation trouvé pour \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 = 3 si l'on désire un intervalle de confiance de 99 %

Note : Les lignes directrices concernant la diffusion des estimations s'appliquent aussi aux intervalles de confiance. Par conséquent, si l'estimation ne peut être diffusée, alors l'intervalle de confiance ne peut être diffusé lui non plus.

10.4 Exemple d'utilisation de tables de c.v. pour obtenir des limites de confiance

Voici la marche à suivre pour calculer un intervalle de confiance de 95 % pour la proportion estimée de personnes qui fument tous les jours parmi celles qui fument (d'après l'exemple 2 de la section 10.2).

$$\hat{X} = .667$$

$$t = 2$$

$\alpha\hat{X} = .011$ est le coefficient de variation de cette estimation selon les tables.

$$CI_X = \{.667 - (2) (.667) (.011), .667 + (2) (.667) (.011)\}$$

$$CI_X = \{.652, .682\}$$

10.5 Comment utiliser les tables de c.v. pour effectuer un test

On peut aussi utiliser les erreurs-types pour effectuer des tests d'hypothèses, une technique qui permet de faire la distinction entre les paramètres d'une population à l'aide d'estimations basées sur un échantillon. Ces estimations peuvent être des nombres, des moyennes, des pourcentages, des rapports, etc. Les tests peuvent être effectués à divers niveaux de signification; 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 basées sur un échantillon pour deux caractéristiques voulues. Supposons aussi que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est σ_d .

Si $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d}$ est compris entre -2 et 2, alors on ne peut tirer aucune conclusion à propos de la différence entre les caractéristiques au niveau de signification de 5 %. Toutefois, si ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0.05.

10.6 Exemple d'utilisation des tables de c.v. pour effectuer un test t

Supposons que nous voulons tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion de personnes qui fument tous les jours, à raison de 10 cigarettes ou plus par jour, ET la proportion de personnes qui fument à l'occasion ou qui ne fument pas du tout mais qui ont déjà fumé tous les jours, à raison de 10 cigarettes ou plus par jour. Dans l'exemple 3 de la section 10.2, nous avons déterminé que l'erreur-type de la différence entre ces deux estimations est égale à 0.014. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d} = \frac{.73 - .48}{.014} = \frac{.25}{.014} = 17.86$$

Puisque $t = 17.86$ est supérieur à 2, on doit conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 0.05.

10.7 Variations ou coefficients de variation réels

Tous les coefficients de variation qui figurent dans les tables de variabilité d'échantillonnage approximative sont effectivement approximatifs, et, donc, non officiels. Toutefois, Statistique Canada peut produire les coefficients de variation réels de variables particulières contre recouvrement des coûts. Les types d'estimation pour lesquels ce service est offert incluent les agrégats, les proportions, les ratios, les différences entre agrégats, proportions ou ratios, ainsi que des types d'analyse plus complexes, comme les estimations des coefficients obtenus par régression linéaire ou par régression logistique. On obtient les coefficients de variation précis grâce à un programme de calcul de la variance réelle fondé sur la «méthode du Jackknife». Cette dernière consiste à diviser les enregistrements du fichier de microdonnées en sous-groupes (ou sous-échantillons) et à déterminer la variation des estimations d'un sous-échantillon à l'autre. Les raisons pour lesquelles un utilisateur pourrait souhaiter connaître la variance réelle sont diverses. En voici quelques-unes.

Premièrement, si un utilisateur désire obtenir des estimations à un niveau géographique inférieur au niveau provincial (par exemple, au niveau urbain ou rural), les tables de variabilité d'échantillonnage approximative publiées ne conviennent pas. Néanmoins, on peut obtenir les coefficients de variation de ce type d'estimations en appliquant la méthode d'estimation «par domaine», au moyen du programme de calcul de la variance réelle.

Deuxièmement, si un utilisateur demande des analyses plus complexes, telles que des estimations de coefficients obtenus par régression linéaire ou par régression logistique,

GUIDE DU FICHIER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

les tables de variabilité d'échantillonnage approximative ne fourniront pas les coefficients de variation associés corrects. Certains progiciels statistiques courants permettent d'incorporer les poids d'échantillonnage aux analyses, mais, souvent, les variances produites ne tiennent pas bien compte de la stratification et de l'agrégation de l'échantillon, contrairement à celles obtenues grâce au programme de calcul de la variance réelle.

Troisièmement, dans le cas de l'estimation de variables quantitatives, il est nécessaire d'utiliser des tables distinctes pour déterminer l'erreur d'échantillonnage. Or, la plupart des variables de l'Enquête nationale sur la santé de la population étant de type nominal, de telles tables n'ont pas été produites. Les utilisateurs qui souhaitent connaître les coefficients de variation de variables quantitatives peuvent néanmoins obtenir ces derniers grâce au programme de calcul de la variance réelle. À noter, toutefois, que le coefficient de variation d'un total quantitatif est généralement plus grand que celui de l'estimation de type nominal correspondante (c'est-à-dire, l'estimation du nombre de personnes qui contribuent à l'estimation quantitative). Si l'estimation de type nominal correspondante ne peut être diffusée, il en sera de même pour l'estimation quantitative. Par exemple, le coefficient de variation de l'estimation du nombre total de cigarettes que fument chaque jour les personnes qui fument tous les jours serait supérieur à celui de l'estimation correspondante du nombre de personnes qui fument tous les jours. Par conséquent, si on ne peut diffuser le coefficient de variation de cette dernière estimation, on ne pourra non plus diffuser celui de l'estimation quantitative correspondante.

Enfin, un utilisateur qui peut se servir des tables de variabilité d'échantillonnage approximative, mais obtient ainsi un coefficient de variation compris dans la fourchette «confidentielle» (de 25,1 % à 33,3 %), ne devrait pas diffuser les estimations associées, à moins que le coefficient de variation soit recalculé grâce au programme de calcul de la variance réelle et qu'on juge que les estimations sont, en fait, diffusables. Cette situation tient au fait que l'estimation des coefficients de variation grâce aux tables de variabilité d'échantillonnage approximative est basée sur une vaste gamme de variables et, donc, jugée grossière, alors que le programme de calcul de la variance réelle produit le coefficient de variation précis associé à la variable en question.

Le programme de calcul de la variance ou du coefficient de variation réel sera offert à partir de novembre 1995 et tout utilisateur intéressé par ce service devrait communiquer avec Diane Stukel au (613) 951-2244, Section des méthodes des statistiques sur la santé, Division des méthodes d'enquêtes-ménages, Statistique Canada. Le temp-machine sera gratuit, mais des frais seront perçus pour les heures de consultation consacrées à l'examen de la demande ainsi qu'à la préparation des passages-machines connexes. Le taux journalier de consultation, fondé sur une journée de 7,5 heures, s'établit à 477,88 \$; il peut être ventilé en un nombre approprié d'heures ou de minutes, au besoin. Évidemment, la durée de la consultation variera d'une demande à l'autre et dépendra de la complexité de l'analyse, du nombre de variables à analyser, etc.

10.8 Seuils pour la diffusion des estimations relatives à l'ENSP

Le nombre minimum d'unités d'échantillonnage (personnes) qui contribuent à un estimateur quelconque aux niveaux provincial, régional et national est indiqué dans le tableau qui suit. Les estimations basées sur des échantillons de taille inférieure au minimum indiqué dans la colonne «Confidentielles» ne peuvent en aucun cas être diffusées.

Table des seuils pour la diffusion des estimations, les provinces et le Canada, membres du ménage, tous les âges

PROVINCE	PUBLIABLES	PUBLIABLES SOUS RÉSERVE	CONFIDENTIELLES
Terre-Neuve	8000	3500	2000
Île-du-Prince-Édouard	2000	1000	500
Nouvelle-Écosse	14500	6500	3500
Nouveau-Brunswick	10000	4500	2500
Québec	52000	22500	13000
Ontario	38500	17000	9500
Manitoba	19000	8500	4500
Saskatchewan	15000	6500	3500
Alberta	30500	13500	7500
Colombie-Britannique	34000	15000	8500
CANADA	38000	16500	9500

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

Table des seuils de diffusion des totaux selon les estimations , les provinces et le Canada, personnes choisies, 12 ans et plus

PROVINCE	PUBLIABLES	PUBLIABLES SOUS RÉSERVE	CONFIDENTIELLES
Terre-Neuve	19500	8500	5000
Île-du-Prince-Édouard	4000	2000	1000
Nouvelle-Écosse	30000	13500	7500
Nouveau-Brunswick	21500	9500	5500
Québec	103000	45500	25500
Ontario	87000	38000	21500
Manitoba	35500	16000	9000
Saskatchewan	28500	13000	7500
Alberta	62000	27500	15500
Colombie-Britannique	73500	32500	18500
CANADA	81500	35500	20000

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

**Table des seuils de diffusion des totaux selon les estimations des groupes d'âges,
Canada, membres du ménage, tous les âges**

GRUPE D'ÂGE	PUBLIABLES SOUS RÉSERVE	PUBLIABLES	CONFIDENTIELLES
0-11	30000	13000	7500
12-24	34000	15000	8500
25-44	33500	14500	8000
45-64	36500	16000	9000
65+	43500	19000	11000

**Table des seuils de diffusion des totaux selon les estimations des groupes d'âges,
Canada, personnes choisies, 12 ans et plus**

GRUPE D'ÂGE	PUBLIABLES SOUS RÉSERVE	PUBLIABLES	CONFIDENTIELLES
12-24	105500	46500	26500
25-44	82500	36000	20500
45-64	73000	32000	18000
65+	48000	21000	12000

11. Facteurs de pondération

Le volet des ménages de l'Enquête nationale sur la santé de la population comporte deux plans de base : un pour les neuf provinces à part le Québec et un autre pour le Québec. Dans les neuf provinces à part le Québec, l'ENSP a fait appel au plan de l'Enquête sur la population active (EPA), avec de nombreuses modifications, pour produire son propre échantillon. Pour cette raison, le calcul des poids est relié à la méthode de pondération employée dans l'EPA. En plus de l'échantillon de l'ENSP tiré du plan de l'EPA, un petit échantillon supplémentaire indépendant a été prélevé par composition aléatoire (CA) dans la région sanitaire de Prince George, en Colombie-Britannique, afin de pouvoir produire des estimations fiables au niveau de la région sanitaire. Au Québec, cependant, on a fait appel à un plan d'échantillonnage à deux degrés; l'échantillon du premier degré a été prélevé en 1992-1993 par l'Enquête sociale et de santé (ESS) et celui du deuxième degré par l'ENSP. Par conséquent, au Québec, le calcul des poids est relié à la méthode de pondération employée dans l'ESS. Pour plus de détails, voir la section intitulée «Plan d'échantillonnage». La méthode des facteurs de pondération pour les provinces sauf le Québec est décrite à la section 11.1 et celle pour le Québec à la section 11.2.

11.1 Méthode de facteurs pondération pour les provinces à part le Québec

Dans les sections qui suivent, on présente d'abord une brève description de la méthode de pondération de base de l'EPA. On décrit ensuite la méthode des facteurs de pondération de base selon la technique de composition aléatoire (CA). Enfin, on donne une description d'un certain nombre d'autres ajustements des poids multiplicatifs qui sont nécessaires pour obtenir les poids finals.

11.1.1 Poids de base de l'EPA

L'EPA fait appel à un plan d'échantillonnage à plusieurs degrés stratifié (il s'agit essentiellement d'un plan à deux degrés mais, dans certains cas, on utilise un plan à trois degrés).

Par exemple, aux endroits où on utilise un plan à deux degrés, le premier degré consiste à sélectionner des grappes selon la méthode systématique d'échantillonnage avec probabilité proportionnelle à la taille sans remise (échantillonnage systématique PPT), la méthode d'échantillonnage systématique aléatoire PPT, ou la méthode des groupes aléatoires. Pour plus de renseignements sur ces méthodes, communiquer avec Diane Stukel, Section des méthodes des statistiques de la santé, Division des méthodes d'enquêtes-ménages, à Statistique Canada. Conformément aux méthodes de prélèvement de l'échantillon mentionnées ci-haut, on calcule ensuite un «poids de la grappe» de l'EPA, qui représente la probabilité inverse de sélection d'une grappe.

La deuxième étape consiste à sélectionner des logements parmi les grappes échantillonnées selon la méthode d'échantillonnage systématique. On calcule ensuite un «poids du logement», qui représente la probabilité inverse de sélection d'un logement, compte tenu que la grappe dont celui-ci fait partie a été sélectionnée.

Un «poids de base de l'EPA» est ensuite attribué en multipliant le poids de la grappe par le poids du logement.

11.1.2 Facteurs de pondération de base de la composition aléatoire

On a recouru à la composition aléatoire (CA) après élimination des banques non valides, de façon à obtenir un échantillon simple stratifié mais aléatoire (sans substitution) des lignes téléphoniques résidentielles. L'échantillon (CA) complet a été tiré de trois strates CA sans lien avec les strates du plan d'échantillonnage de l'EPA utilisé dans la région de Prince George. Le «facteur de pondération de base CA» correspond donc à la probabilité inverse de sélectionner une ligne de téléphone résidentielle à partir d'une banque valide.

11.1.3 Corrections supplémentaires apportées aux facteurs de pondération de base

Toutes les corrections qui suivent, apportées aux facteurs de pondération de base de l'EPA, ont pour but de compenser les particularités du plan d'échantillonnage de l'EPA. Certaines s'appliquent néanmoins aux facteurs de pondération de base CA, étant donné leur association à la méthode de composition aléatoire (notamment la Correction pour les logements multiples et la Correction pour la non-réponse du ménage).

Correction 1 : Correction pour le groupe de renouvellement

L'échantillon de l'EPA se compose de 6 «groupes de renouvellement» (quoique dans certaines régions éloignées et pour la strate des appartements, ce nombre puisse être différent). Dans l'EPA, on retire le groupe de renouvellement utilisé depuis six mois de l'échantillon et on le remplace par un nouveau groupe. On procède ainsi pour alléger le fardeau des répondants. Pour l'ENSP, on prélève un échantillon de l'EPA en choisissant un nombre entier (de 1 à 6) correspondant au groupe de renouvellement. Il arrive cependant qu'on prenne un nombre fractionnaire pour satisfaire l'exigence de la taille de l'échantillon. Ainsi, l'ENSP pourrait nécessiter 2,3 renouvellements. On en réclamera donc trois de l'EPA et on «supprimera» subséquemment les sept dixièmes de renouvellement inutiles (voir Correction pour la stabilisation). Par conséquent, la première correction effectuée pour compenser l'effet du nombre entier est la suivante :

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

nombre de renouvellements dans une strate de l'EPA utilisé pour l'EPA (habituellement 6)

nombre entier de renouvellements dans une strate de l'EPA réclamé pour l'ENSP

Dans deux cas, à Winnipeg, et dans un autre, à Vancouver, on a dû combiner trois strates de l'EPA avant de sélectionner des grappes. On a donc procédé à une correction supplémentaire des facteurs de pondération de base de l'EPA pour tenir compte du fait que les grappes avaient été sélectionnées par échantillonnage systématique PPT (randomisé) du groupe de strates plutôt que des strates habituelles de l'EPA.

Correction 2 : Correction pour l'expansion de la grappe

Il se peut que la taille de la grappe augmente entre le recensement et le moment où elle apparaît sur la liste de l'EPA. La probabilité de sélection de la grappe repose sur les données du recensement, qui pourraient être périmées. Il s'ensuit que le nombre de logements de l'échantillon de l'EPA augmente très légèrement avec une croissance modérée du parc immobilier. On peut tolérer une croissance sensible du nombre d'habitations dans une grappe isolée avant que l'échantillon supplémentaire ne pose un problème de collecte sur le terrain. Toutefois, si cette croissance touche plusieurs grappes attribuées à un intervieweur, la charge de travail de ce dernier peut s'en trouver alourdie. Quand les grappes subissent une forte croissance, on procède habituellement à un sous-échantillonnage afin de ne pas surcharger l'intervieweur. L'ENSP permet un sous-échantillonnage analogue des grappes qui ont enregistré une croissance modérée, même si cette dernière n'est pas suffisante pour qu'on en tienne compte dans l'EPA. Le second facteur de correction multiplicatif correspond donc à l'application du rapport de sous-échantillonnage inverse aux grappes sous-échantillonnées dans le cadre de l'ENSP.

Correction 3 : Correction pour la stabilisation

La stabilisation permet de plafonner la taille de l'échantillon dans un secteur de stabilisation, donc d'éviter que les frais connexes deviennent prohibitifs. Par «secteur de stabilisation», on entend les grappes de la base de sondage des ménages à revenu élevé et des ménages vivant en appartement, consistant du groupement des strates de la base de sondage habituelle. La «stabilisation» résout le problème que pourrait poser la croissance d'un secteur de stabilisation. Une telle croissance est assez importante pour devenir préoccupante même après correction pour l'expansion de la grappe, bien qu'aucune grappe ne se soit suffisamment développée pour qu'on la considère à la racine du mal. On remédie au problème en sous-échantillonnant le secteur de stabilisation. Outre la stabilisation ordinaire, c'est à ce niveau qu'on «supprime» la fraction du groupe de

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

renouvellement de l'EPA dont l'ENSP n'a pas besoin par sous-échantillonnage (voir Correction pour le groupe de renouvellement). La troisième correction s'effectue comme suit :

nombre de logements dans une grappe sélectionnés pour l'EPA

nombre de logements dans une grappe utilisés pour l'ENSP

Correction 4 : Correction pour les logements multiples

Il arrive que l'intervieweur découvre une unité d'habitation comptant deux ou plusieurs logements dans une liste qu'on croyait ne comprendre que des habitations unifamiliales. Pareille situation peut survenir, par exemple, quand un appartement en sous-sol fait partie d'une habitation mais possède sa propre entrée. Puisqu'on interroge les occupants d'un seul des deux logements (choisi au hasard), le facteur de pondération associé au logement sélectionné est doublé. La quatrième correction se rapporte donc au nombre de logements privés occupés que compte vraiment la liste utilisée pour l'enquête. Dans la plupart des cas, le facteur de correction multiplicatif est égal à 1.

L'échantillon CA fait l'objet d'une correction analogue. En effet, un ménage qui possède, par exemple, deux lignes téléphoniques résidentielles, apparaissant dans la base de sondage CA, aura deux fois plus de chances d'être sélectionné qu'avec une seule ligne. On applique donc un facteur de correction multiplicatif égal à l'inverse du nombre de lignes téléphoniques résidentielles recensées aux ménages retenus pour lesquels l'existence de plusieurs lignes de téléphone a été confirmée. Notons que cette correction est l'inverse de celle effectuée pour les logements multiples dans l'EPA.

Correction 5 : Correction pour la non-réponse du ménage

Malgré tous ses efforts, il est inévitable que l'intervieweur n'obtienne pas de réponse de certains ménages pour l'une des raisons suivantes : refus, circonstances particulières, obstacle linguistique, logement vide, absence temporaire ou problème d'ordinateur. On compense les non-réponses en corrigeant proportionnellement les facteurs de pondération des ménages répondants. La cinquième correction s'effectue comme suit :

somme des facteurs de pondération des ménages échantillonnés pour le groupe strate-saison de l'ENSP

somme des facteurs de pondération des ménages répondants pour le groupe strate-saison de l'ENSP.

Notons qu'on procède à cette correction chaque saison, au niveau de la strate de l'ENSP, et que chaque strate correspond à des groupes de strates de l'EPA. On a retenu ce niveau car il s'agit du plus petit palier géographique garantissant la stabilité (à savoir, les corrections sont presque toujours inférieures ou égales à 2,5). La correction a été calculée séparément pour chaque saison car le taux de non-réponse varie significativement d'une saison à l'autre. Dans le cas qui nous intéresse, les deux premiers trimestres de collecte figurent la «saison» estivale, alors que les deux derniers représentent la «saison» hivernale. Dans les rares cas où la correction pour les non-réponses dépasse 2,5, on a procédé à un second calcul au niveau de la strate de l'ENSP plutôt que du groupe strate-saison. Les «corrections» précitées correspondent aux facteurs de pondération de base de l'EPA multipliés par les facteurs de correction établis jusqu'à présent (à savoir, les corrections 1 à 4). La correction repose sur l'hypothèse que les ménages interrogés présentent les mêmes caractéristiques que ceux qui auraient dû l'être. Si l'hypothèse ne se vérifie pas, les estimations seront légèrement faussées. Précisons qu'on a pu rejoindre par la suite certaines personnes absentes lors de la période de collecte établie. Ces cas ont été traités comme si les réponses avaient été recueillies durant la période initialement prévue. Étant donné leur nombre, ces cas ne devraient pas avoir d'influence significative sur les données saisonnières.

Puisque l'échantillon CA soulève aussi des problèmes de non-réponse, une correction distincte s'impose, bien que la méthode utilisée soit pratiquement identique à celle décrite précédemment. La première distinction est que la correction s'effectue au niveau de la strate CA et de la saison. L'autre est que les «facteurs de pondération» mentionnés dans la formule correspondent maintenant aux facteurs de pondération de base CA multipliés par le facteur de correction pour les logements multiples de la composition aléatoire.

Correction 6 : Correction pour la méthode d'exclusion

Tel qu'indiqué dans la partie décrivant le plan d'échantillonnage, on a écarté ou rejeté une partie des ménages échantillonnés lors des deux derniers trimestres de collecte des données après avoir déterminé que ces ménages ne comptaient pas de jeunes ni d'enfants (à savoir, une personne de moins de 25 ans). Les ménages «rejetés» viennent de la partie de l'échantillon «enfants» épurée d'après la composition du ménage. On a recouru à cette méthode pour compenser la surreprésentation des ménages de petite taille dans l'échantillon et la sous-représentation des ménages de grande taille. Le second type de ménage comprend plus souvent des parents et leurs enfants, alors que le premier est surtout constitué de célibataires, de personnes âgées ou de couples sans enfants.

Puisqu'on en écarte ou «rejette» certains, les ménages sans jeunes ni enfants se retrouvent uniquement dans l'échantillon «adultes» et dans la partie de l'échantillon «enfants» qui n'a pas été épurée. Pour compenser l'exclusion d'une partie de l'échantillon, on applique un facteur de correction multiplicatif aux facteurs de pondération des ménages sans jeunes ni enfants de l'échantillon «adultes» et de la partie non épurée de l'échantillon «enfants». Pour la sixième correction, on multiplie les facteurs de pondération par l'inverse de un moins le taux d'épuration global pour la grappe. À l'Île-du-Prince-Édouard, la correction était légèrement différente car la méthode d'exclusion, notamment, a été appliquée aux quatre périodes de collecte des données plutôt qu'aux deux derniers trimestres. Remarquons aussi qu'on n'a pas appliqué cette correction à la strate des ménages vivant en appartement, à celle des ménages à revenu élevé ni à celle des ménages habitant dans une région éloignée, la méthode d'exclusion n'ayant pas été employée avec ces trois strates.

11.1.4 Corrections supplémentaires pour les membres du ménage

Les données du questionnaire du ménage concernent tous les membres du ménage échantillonné. Voici comment on pondère chaque sujet du fichier. Tout d'abord, on multiplie le facteur de pondération de base de l'EPA par les facteurs de correction 1 à 6 et le facteur de correction pour la non-réponse du membre du ménage donné plus bas, de manière à obtenir un coefficient de pondération «intermédiaire» pour l'EPA. Ensuite, on multiplie le facteur de pondération de base CA par le facteur de correction pour les logements multiples, le facteur de correction pour la non-réponse du ménage et le facteur de correction pour la non-réponse du membre du ménage, ce qui donne le facteur de pondération «intermédiaire» CA. Cela fait, on regroupe les deux échantillons indépendants (EPA et CA) de Prince George grâce à la «correction pour le membre du ménage en vue de la combinaison des échantillons EPA et CA», tel qu'indiqué ci-dessous, au moyen des facteurs de pondération «intermédiaires» pour l'EPA et la composition aléatoire. Enfin, on procède à une analyse comparative de l'échantillon combiné de Prince George et de l'échantillon habituel de l'EPA provenant d'un autre endroit.

Correction 7A : Correction pour la non-réponse du membre du ménage

Il peut arriver qu'on possède des renseignements incomplets sur certains membres du ménage, même si ce dernier fait partie des ménages «répondants». Les membres du ménage dans cette situation sont considérés comme des non-répondants, si bien qu'on doit corriger les données des répondants du même groupe d'âge et du même sexe de la province. Cette correction s'effectue comme suit :

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

somme des facteurs de pondération des membres du ménage échantillonné pour le groupe province-âge-sexe

somme des facteurs de pondération des répondants du ménage pour le groupe province-âge-sexe

Les «facteurs de pondération» précités correspondent aux facteurs de pondération de base de l'EPA multipliés par les facteurs de correction établis jusqu'à présent (à savoir, corrections 1 à 6). Les groupes d'âge utilisés pour les deux sexes sont les suivants : 0-11, 12-24, 25-44, 45-64 et 65 et plus.

Le facteur de correction pour la non-réponse du membre du ménage de l'échantillon CA est appliqué séparément bien que la correction en elle-même soit identique à la précédente. La première distinction est que les «facteurs de pondération» mentionnés dans la formule correspondent aux facteurs de pondération de base CA, multipliés par les facteurs de correction 4 et 5. La seconde est que la correction s'applique aux groupes âge-sexe de Prince George plutôt qu'aux groupes âge-sexe de chaque province.

Correction 8A : Correction pour le membre du ménage en vue de la combinaison des échantillons EPA et CA

Quoiqu'on ait prélevé un échantillon distinct par la technique de la composition aléatoire dans la région de Prince George, en Colombie-Britannique, cette région a également fait l'objet d'un échantillonnage ordinaire dans le cadre de l'EPA. Les deux échantillons étant indépendants mais les régions se chevauchant, on a recouru à une double base de sondage pour les agréger. L'approche retenue a été appliquée à l'ensemble de la Colombie-Britannique. La méthode consistait essentiellement à trois corrections des facteurs de pondération. La première correction s'applique aux «facteurs de pondération CA» de l'échantillon CA provenant des trois strates échantillonnées par composition aléatoire. La deuxième concerne les «facteurs de pondération de l'EPA» de l'échantillon de l'EPA que recourent les trois strates CA. Enfin, la dernière touche les «facteurs de pondération de l'EPA» de la partie de l'échantillon de l'EPA extérieure aux trois strates CA, mais faisant partie de la Colombie-Britannique. Les facteurs de correction sont passablement complexes; pour plus de précisions à leur sujet, on lira Skinner, C. et Rao, J.N.K. (JASA, 1996). Les «facteurs de pondération CA» mentionnés plus haut correspondent aux facteurs de pondération de base CA multipliés par les facteurs de correction 4, 5 et 7A. Les «facteurs de pondération de l'EPA» sont les facteurs de pondération de base de l'EPA, multipliés par les facteurs de correction 1 à 6 et 7A.

Correction 9A : Correction pour l'analyse comparative des membres du ménage

On peut obtenir des estimations indépendantes tous les mois sous forme de projections démographiques pour certains groupes d'âge des deux sexes, par province. Ces projections reposent sur les données les plus récentes du recensement, les registres de naissances et de décès et les estimations de migration. Lors de l'étape ultime, ces renseignements complémentaires permettent de calculer les facteurs de pondération finals. La correction pour l'analyse comparative des membres du ménage fait en sorte que la somme des facteurs de pondération finals correspond aux projections démographiques pour les variables auxiliaires concernées, c'est-à-dire les catégories d'âge suivantes pour les deux sexes : 0-11, 12-24, 25-44, 45-64, 65 et plus. On procède à la correction comme suit:

projections démographiques pour un groupe province-âge-sexe

somme des facteurs de pondération des répondants d'un ménage pour un groupe province-âge-sexe

Les «facteurs de pondération» dont il est question dans la formule ci-dessus comprennent les facteurs de pondération obtenus jusqu'à présent, soit, pour l'EPA, les facteurs de pondération de base EPA multipliés par les huit derniers facteurs de correction et, pour la composition aléatoire, les facteurs de pondération de base CA multipliés par les facteurs de corrections 4, 5, 7A et 8A. Les données ayant été recueillies durant les quatre trimestres, les projections démographiques servant à l'analyse comparative sont les projections moyennes pour les quatre trimestres durant lesquels s'est déroulée l'enquête. La dernière intervention consiste à multiplier les «facteurs de pondération» obtenus jusqu'à présent par le facteur de correction pour l'analyse comparative des membres du ménage.

11.1.5 Corrections supplémentaires pour les membres du panel

Les données du questionnaire du membre du panel viennent d'une seule personne d'au moins 12 ans du ménage échantillonné. Le facteur de pondération final associé à chaque sujet du fichier est calculé de la façon suivante. On multiplie les facteurs de pondération de base de l'EPA par les facteurs de correction 1 à 6 et 7B à 10B décrits plus bas afin d'obtenir un facteur de pondération «intermédiaire» pour l'EPA. Ensuite, on multiplie le facteur de pondération de base CA par les facteurs de correction 4, 5 et 8B à 10B décrits plus bas afin d'obtenir un facteur de pondération «intermédiaire» CA. Les deux échantillons indépendants (EPA et CA) de la région de Prince George sont subséquentement combinés par application du facteur de correction pour les membres du panel, tel qu'indiqué ci-dessous, à

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

partir des facteurs de pondération «intermédiaires» pour l'EPA et la composition aléatoire. Enfin, on effectue une analyse comparative de l'échantillon combiné de Prince George et d'un échantillon ordinaire de l'EPA d'une autre région.

Correction 7B : Correction pour l'intégration de l'ELNE

Au cours des deux derniers trimestres, l'ENSP sélectionne des répondants éventuels pour le questionnaire du membre du panel tant de l'ENSP que de l'ELNE. Dans certains ménages, on choisit jusqu'à 4 enfants (de moins de 12 ans), qui répondront aux questions de l'ELNE. Les données recueillies ne font pas partie du fichier de microdonnées actuel. Dans d'autres ménages, on retient un membre de 12 ans et plus et on le prie de répondre au questionnaire de l'ENSP. Les données résultantes se trouvent dans le fichier de microdonnées existant. Pour plus de renseignements sur l'intégration de l'ELNE, voir la partie sur le plan d'échantillonnage. Les répondants de 12 ans et plus des ménages avec enfants viennent de l'échantillon «adultes». Puisque les ménages avec enfants de l'échantillon «enfants» ne contribuent pas aux estimations relatives aux personnes de 12 ans et plus, on corrige les facteurs de pondération applicables aux membres des ménages avec enfants de l'échantillon «adultes» en les multipliant par la proportion inverse de l'échantillon total que représente l'échantillon «adultes». La correction s'effectue au niveau de la grappe pour les personnes de plus de 12 ans. Pour celles de 12 ans, une correction distincte s'impose aux groupes de strates de l'EPA (correspondant habituellement aux strates de l'ENSP), correction cohérente avec la correction 9B, également effectuée à ce niveau.

Correction 8B : Probabilité de sélection inverse des membres du panel

Tel qu'indiqué précédemment, un membre de 12 ans et plus de chaque ménage échantillonné devient membre du panel. Cette sélection exige une correction. C'est pourquoi on recourt à la probabilité de sélection inverse. L'intention initiale était de donner à chaque membre du ménage de 12 ans et plus la même chance d'être sélectionné, chance correspondant à l'inverse du nombre de personnes de 12 ans et plus que compte le ménage. Malheureusement, une erreur dans l'IPAO a interdit la sélection des enfants de 12 ans au cours du premier trimestre. En compensation, au lieu d'appliquer la même probabilité d'être sélectionné à chaque membre du ménage durant les deux derniers trimestres, on a relevé la probabilité de sélection des enfants de 12 ans. Ainsi, à l'Île-du-Prince-Édouard, les enfants de 12 ans avaient deux fois plus de chances d'être retenus qu'un autre membre du ménage de 13 ans et plus. La probabilité était de 1,75 dans les autres régions du Canada.

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

On a également calculé la probabilité de sélection inverse pour l'échantillon CA.

Correction 9B : Correction pour les enfants de 12 ans

À cause de l'erreur précitée, les enfants de 12 ans n'ont pu être sélectionnés qu'au cours des deux dernières périodes de la collecte des données. Afin que ce groupe de la population soit correctement représenté, on a corrigé les facteurs de pondération de façon à tenir compte du premier trimestre, où les probabilités d'être retenu étaient nulles. Cette correction s'est appliquée aux groupes de strates de l'EPA qui correspondent habituellement aux strates de l'ENSP, outre celles des ménages habitant dans une région éloignée et des ménages à revenu élevé. Dans les ménages avec enfants, les enfants de 12 ans auraient pu venir de l'échantillon «adultes», peu importe le trimestre, mais en réalité ils n'ont été sélectionnés qu'au cours des deux derniers. Puisqu'on a réuni 40 % de l'échantillon «adultes» au cours des deux derniers trimestres pour la plupart des strates de l'ENSP, on a relevé le facteur de pondération des enfants de 12 ans sélectionnés durant ce laps de temps en le multipliant par l'inverse du taux, soit 2,5. Pareillement, dans les ménages avec jeunes mais sans enfants, les enfants de 12 ans pouvaient être sélectionnés de l'échantillon «adultes» ou de l'échantillon «enfants». Au cours des deux premiers trimestres, la sélection ne s'est cependant pas effectuée dans l'échantillon «adultes», contrairement à ce qui aurait dû se produire. C'est pourquoi le facteur de pondération des enfants de 12 ans dans les ménages avec jeunes mais sans enfants a été multiplié par le rapport entre le pourcentage de l'échantillon total de la strate de l'ENSP où la sélection aurait dû se produire et le pourcentage de l'échantillon total où la sélection a effectivement eu lieu, soit par 1,6. Enfin, les enfants de 12 ans n'auraient pu être sélectionnés dans les ménages sans jeunes ni enfants, si bien qu'on n'a apporté aucune correction aux facteurs de pondération pour les enfants de 12 ans dans ce type de ménage. Soulignons que les taux diffèrent légèrement pour l'Île-du-Prince-Édouard, la strate des ménages vivant en appartement, celle des ménages à revenu élevé et celle des ménages habitant dans une région éloignée.

Cette correction a été appliquée aux échantillons EPA et CA.

Correction 10B : Correction pour la non-réponse du membre du panel

Il arrive qu'on ne possède pas tous les renseignements voulus sur le membre du ménage sélectionné pour faire partie du panel, même si le ménage compte parmi les «répondants». Dans ce cas, le membre du panel est considéré comme un non-répondant, de sorte qu'on doit procéder à une correction pour les membres du panel du même groupe âge-sexe-province faisant partie des répondants.

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

Cette correction s'effectue comme suit :

somme des facteurs de pondération des membres du panel pour le groupe province-âge-sexe

somme des facteurs de pondération des répondants du panel pour le groupe province-âge-sexe

Les «facteurs de pondération» dont il est question ci-dessus correspondent aux facteurs de pondération de base de l'EPA multipliés par les facteurs de correction établis jusqu'à présent (à savoir, 1 à 6 et 7B à 9B). Les groupes d'âge des deux sexes sont les suivants : 12-24, 25-44, 45-64 et 65 et plus, puisque seuls les sujets de 12 ans et plus répondent au questionnaire du membre du panel.

La correction pour la non-réponse du membre du panel de l'échantillon CA s'effectue à part, bien qu'elle soit pratiquement identique à la précédente. La première distinction est que les «facteurs de pondération» de la formule désignent les facteurs de pondération de base CA, multipliés par les facteurs de correction 4, 5, 8B et 9B. La seconde est qu'on procède à la correction pour les groupes âge-sexe de Prince George plutôt que les catégories correspondantes de chaque province.

Correction 11B : Correction pour le membre du panel en vue de la combinaison des échantillons EPA et CA

Cette correction ressemble à celle effectuée pour les membres du ménage, à la différence que les «facteurs de pondération CA» dont il est question sont les facteurs de pondération de base CA, multipliés par les facteurs de corrections 4, 6, 9B, 10B et 11B. Les «facteurs de pondération de l'EPA» quant à eux correspondent aux facteurs de pondération de base de l'EPA, multipliés par les facteurs de correction 1 à 6 et les quatre premiers facteurs de correction de la présente partie.

Correction 12B : Correction pour l'analyse comparative des membres du panel

Cette correction est analogue à celle s'appliquant aux membres du ménage et prend la forme suivante :

projections démographiques pour le groupe province-âge-sexe

somme des facteurs de pondération pour les répondants du panel du groupe province-âge-sexe

Les groupes d'âge utilisés pour les deux sexes étaient 12-24, 25-44, 45-64 et 65 et plus. Les «facteurs de pondération» qui apparaissent dans l'équation précédente

GUIDE DU FICHIER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

correspondent aux facteurs de pondération établis jusqu'à présent, c'est-à-dire, pour l'EPA, les facteurs de pondération de base EPA multipliés par les facteurs de correction 1 à 6 et les cinq premiers facteurs de correction de la présente partie et, pour la composition aléatoire, les facteurs de pondération de base CA multipliés par les facteurs de correction 4, 6, et 9B à 12B. Puisque les données ont été recueillies pendant quatre trimestres, les projections démographiques utilisées pour l'analyse comparative correspondent aux projections moyennes pour les quatre trimestres qu'a duré l'enquête. Enfin, on obtient les facteurs de pondération finals en multipliant les «facteurs de pondération» établis jusqu'à présent par le facteur de correction pour l'analyse comparative des membres du panel qui précède.

11.2 Méthode de pondération pour le Québec

Le plan d'échantillonnage de l'Enquête nationale sur la santé de la population intégrait un sous-échantillon de l'*Enquête sociale et de santé* (ESS) (voir la partie sur le plan d'échantillonnage pour plus de précisions). C'est pourquoi le calcul des facteurs de pondération de l'ENSP est lié aux méthodes utilisées dans le cadre de l'ESS. Les parties qui suivent décrivent les méthodes de pondération de l'ESS et la manière dont on s'y est pris pour obtenir les facteurs de pondération applicables aux participants de l'ENSP.

11.2.1 Facteurs de pondération de l'ESS

On a déterminé l'apport de l'ESS aux facteurs de pondération comme suit :

Facteurs de pondération pour la grappe de l'ESS

L'ESS recourait à un plan d'échantillonnage stratifié à plusieurs degrés. Après stratification, on a sélectionné des grappes de chaque strate avec probabilité proportionnelle à la taille (PPT). La taille retenue était le nombre de ménages dans la grappe, selon le Recensement de 1986. Le «**facteur de pondération pour la grappe de l'ESS**» correspond à la probabilité inverse de sélectionner une grappe.

Facteurs de pondération pour le logement de l'ESS

Après sélection d'une grappe, on a déterminé le nombre de logements qui seraient choisis dans la grappe. Chaque logement avait autant de chances d'être sélectionné qu'un autre. Le «**facteur de pondération pour le logement de l'ESS**» correspond donc à la probabilité inverse de sélectionner le logement dans la grappe, multipliée par le facteur de pondération pour la grappe de l'ESS.

11.2.2 Facteurs de pondération de base pour les logements de l'ENSP

L'échantillon de l'ENSP a été constitué en deux grandes étapes. Tout d'abord, on a déterminé le sous-ensemble de grappes de l'ESS qui servirait à l'ENSP. Ensuite, il a fallu prélever un sous-échantillon de logements dans chaque grappe de l'ESS retenue.

Probabilité de rétention d'une grappe de l'ESS pour l'ENSP

Les strates de l'ESS étaient parfois très restreintes, si bien que celles de l'ENSP pouvaient en inclure plusieurs. Par ailleurs, on devait retenir un nombre fixe de

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

grappes de chaque strate de l'ENSP. Quand cette strate comprenait plusieurs strates de l'ESS, la répartition des grappes entre les strates de l'ESS était proportionnelle au nombre de ménages dans chaque strate de l'ESS. Cette méthode a permis d'obtenir un échantillon de grappes PPT pour chaque strate de l'ENSP. Les échantillons fractionnaires ont été arrondis au hasard à l'entier supérieur ou inférieur. Une fois le nombre de grappes de la strate de l'ESS déterminé, chaque grappe de la strate de l'ESS avait la même probabilité d'être conservée dans la plupart des cas. Les seules exceptions concernent les grappes où le nombre de logements avait augmenté de plus de 150 % entre le Recensement de 1986 et le listage de l'ESS de 1992-1993. Ces grappes ont fait l'objet d'une plus grande probabilité de sélection (relèvement de 100 % ou de 40 % de la probabilité de rétention).

Probabilité de rétention d'un logement de l'ESS pour l'ENSP

Dans les grappes retenues, seuls les logements choisis pour l'ESS pouvaient être sélectionnés en vue de l'ENSP. Les logements débordant du cadre de l'ESS (entreprises, institutions, habitations démolies ou logements vacants) avaient une probabilité de rétention égale à un. À partir des logements couverts par l'ESS, on a d'abord gardé un nombre fixe de logements de chaque grappe pour l'ENSP. Un autre sous-groupe de logements a été écarté en raison de la composition du ménage de l'ESS. La probabilité qu'on garde un logement en raison de la composition du ménage apparaît au tableau suivant.

Probabilité de rétention du logement sélectionné au départ pour l'ENSP

Composition du ménage de l'ESS	Probabilité de rétention
Ménage avec enfants (moins de 12 ans)	1
Ménages d'un seul membre	1/3
Ménage comptant au moins un jeune (12 à 24 ans)	5/6
Autres ménages	1/2

Le «**facteur de pondération de base pour le logement**» correspond au facteur de pondération pour le logement de l'ESS, multiplié par le produit inverse de la probabilité de rétention de la grappe de l'ESS avec la probabilité de rétention du logement de l'ESS. La probabilité de rétention de logement de l'ESS inclut la

probabilité que le logement soit sélectionné pour l'ENSP au départ et la probabilité qu'il soit gardé en raison de la composition du ménage.

11.2.3 Corrections supplémentaires aux facteurs de pondération de base

Correction pour les logements multiples

Il arrive qu'un intervieweur visitant une unité d'habitation découvre un logement supplémentaire qui a échappé au listage des grappes. Un appartement au sous-sol en est un bon exemple. Dans un tel cas, chaque logement devient un multiple. On sélectionne un logement au hasard et on interroge l'occupant. Le facteur de pondération du logement choisi est ensuite corrigé par un facteur multiplicatif égal au nombre de multiples.

Correction pour l'expansion de la grappe

Dans certains cas, l'ENSP a nécessité le relistage des grappes. Quand la comparaison des chiffres de l'ESS et de l'ENSP révélait une hausse de 15 à 30 %, on procédait à la correction suivante :

numération de l'ENSP

numération de l'ESS

Cette correction a été appliquée à chaque logement sélectionné de la grappe. On a supposé que les écarts inférieurs à 15 % étaient négligeables, auquel cas le facteur de correction a été ramené à un. Pour tous les logements concernés, on multiplie les facteurs de correction pour les logements multiples et pour l'expansion de la grappe par le facteur de pondération de base pour le logement de manière à obtenir le «**facteur de pondération préliminaire**».

Quand l'expansion de la grappe était supérieure à 30 %, des logements supplémentaires ont été sélectionnés parmi les logements additionnels de la grappe pour l'ENSP. Le «**facteur de pondération préliminaire**» des logements supplémentaires retenus est égal au produit inverse de la probabilité de sélection de la grappe de l'ESS et de la probabilité de rétention de la grappe pour l'ENSP multiplié par le facteur de correction pour les logements multiples et le facteur de correction suivant :

nombre de logements supplémentaires listés

nombre de logements supplémentaires sélectionnés

Puisque aucun de ces logements n'a conduit à une entrevue dans le cadre de l'ESS, il est impossible de les répartir entre les différentes catégories de l'ESS relatives à la composition du ménage.

Correction pour la non-réponse du ménage

Un certain nombre de non-réponses est inévitable dans une enquête. L'ENSP ne fait pas exception à la règle. Pour tenir compte du nombre total de ménages non répondants, on a procédé à la correction suivante :

somme des facteurs de pondération des ménages répondants et non répondants

somme des facteurs de pondération des ménages non répondants

Dans ce cas, les facteurs de pondération correspondent aux facteurs de pondération préliminaires.

Une autre correction s'impose au niveau du secteur de pondération pour les non-réponses. En ce qui concerne les logements couverts par l'ESS, le secteur de pondération pour les non-réponses correspond à l'intersection entre une strate de l'ENSP et un type de ménage de l'ESS (les quatre types décrits au tableau 11.1), chaque trimestre. Si le facteur de correction résultant est trop élevé (supérieur à 2,5), les types de ménages sont systématiquement regroupés jusqu'à ce qu'on obtienne un facteur inférieur à 2,5.

Dans le cas des logements ajoutés à la suite d'une expansion de la grappe supérieure à 30 % lors du relistage de l'ENSP, le secteur de pondération correspond aux logements supplémentaires de la grappe, par trimestre.

Les logements qui débordaient du cadre de l'ESS sont regroupés en deux secteurs de pondération par trimestre afin de faciliter la correction pour la non-réponse. Le premier groupe rassemble les logements dont le code-réponse de l'ESS était 10 (démoli, vide, abandonné). Le second comprend les logements dont le code-réponse de l'ESS était 18 (institution ou entreprise).

On obtient le «**facteur de pondération démographique**» en multipliant le facteur de pondération préliminaire par le facteur de correction pour la non-réponse du ménage.

11.2.4 Corrections supplémentaires pour les membres du ménage

Tous les membres du ménage sélectionné doivent répondre au questionnaire du ménage. La non-réponse ne survient que si certains membres du ménage seulement répondent aux questions. On doit alors procéder à une correction pour tenir compte des non-répondants (leur nombre était négligeable dans le cas présent). La dernière étape consiste en une analyse comparative des facteurs de pondération d'après les projections démographiques pour l'âge et le sexe.

Correction pour la non-réponse du membre du ménage

Cette correction permet de tenir compte des membres d'un ménage répondant (c'est-à-dire d'un ménage qui a accepté de répondre au questionnaire démographique) qui refusent de répondre au questionnaire personnel. La correction est la suivante :

somme des facteurs de pondération des répondants et des non-répondants du groupe âge-sexe

somme des facteurs de pondération des répondants du groupe âge-sexe

Le facteur de pondération mentionné dans la formule est le facteur de pondération démographique. Les groupes d'âge pour les deux sexes sont 0-11, 12-24, 25-44, 45-64, 65 et plus. Chaque groupe âge-sexe de la province fait l'objet d'une correction.

Correction pour l'analyse comparative des membres du ménage

Il existe des projections démographiques provinciales pour différents groupes âge-sexe. Les facteurs de pondération des répondants d'un groupe âge-sexe donné sont corrigés afin que leur somme soit égale aux projections démographiques du même groupe de la façon suivante :

projections démographiques du groupe âge-sexe

somme des facteurs de pondération des répondants du groupe âge-sexe

Le facteur de pondération utilisé ici correspond au facteur de pondération démographique multiplié par le facteur de correction pour la non-réponse du ménage. Voici les groupes d'âge pour les hommes et les femmes : 0-11, 12-24, 25-44, 45-64, 65 et plus. Notons que trois régions sanitaires du Nord étaient exclues de l'ESS, donc de l'ENSP. Ces régions représentent environ 0,5 % de la population visée par l'ENSP. Les projections démographiques ont été corrigées pour en tenir compte. L'enquête ayant duré quatre trimestres, les projections démographiques servant à l'analyse

comparative correspondent aux projections moyennes pour les quatre trimestres durant lesquels s'est déroulée l'enquête.

On obtient le «**facteur de pondération pour le membre du ménage**» en multipliant le facteur de pondération démographique par le facteur de correction pour la non-réponse du membre du ménage et le facteur de correction pour l'analyse comparative des membres du ménage.

11.2.5 Corrections supplémentaires pour les membres du panel

Un membre de chaque ménage répondant a été sélectionné pour devenir membre du panel de l'étude longitudinale. S'il s'agissait d'un enfant de moins de 12 ans vivant dans un logement de l'échantillon «enfants» (voir la partie sur le plan d'échantillonnage pour la définition de l'échantillon «enfants»), tous les enfants du ménage, jusqu'à concurrence de quatre, étaient priés de répondre au questionnaire de l'Enquête longitudinale nationale sur les enfants (ELNE). Dans les autres cas, on posait au membre du panel (de 12 ans et plus) la série de questions supplémentaires de l'ENSP. Plusieurs corrections ont dû être apportées en raison du plan d'échantillonnage et des non-réponses à ce questionnaire.

Correction pour l'intégration de l'ELNE

Quand le ménage de l'échantillon «enfants» compte au moins un enfant, un de ceux-ci est désigné pour devenir membre du panel de l'ENSP. L'enfant en question et les autres enfants du ménage, jusqu'à concurrence de quatre, répondent au questionnaire du membre du panel de l'ENSP. Ces données ne se retrouvent pas dans la banque de microdonnées actuelle. Il faut cependant effectuer une correction pour tenir compte des adultes et des jeunes vivant dans ces logements qui n'ont pas eu la possibilité d'être sélectionnés. La correction ne s'applique qu'aux adultes et aux jeunes sélectionnés pour le panel de l'étude longitudinale dans l'échantillon «adultes» où l'ENSP a permis la découverte d'enfants.

La correction correspond à l'inverse du taux de sous-échantillonnage pour l'échantillon «adultes» et dépend des catégories dans lesquelles tombe le logement, parmi les suivantes :

- 1) L'ESS mentionnait-elle la présence d'enfants dans le logement ? (oui ou non)
- 2) À quelle catégorie de densité urbaine de l'ESS le logement appartenait-il ?

Les logements où l'ESS a permis la découverte d'enfants et les logements sans enfants ont dû faire l'objet d'une correction distincte, car le taux de sous-échantillonnage était différent dans les deux cas. Pour Montréal et les capitales régionales, la correction

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

a été apportée au niveau de la grappe tandis que pour les plus petites agglomérations urbaines et les régions rurales, elle a été effectuée au niveau de la strate de l'ENSP. Il existe une seule exception à cette règle; elle est expliquée plus loin à la partie concernant la correction pour les enfants de 12 ans.

Probabilité de sélection inverse des membres du panel

Dans les logements de l'échantillon «enfants» où on n'a pas recensé d'enfants de moins de 12 ans et dans les logements de l'échantillon «adultes», chaque membre du ménage de 12 ans et plus devait avoir au départ la même chance d'être sélectionné pour faire partie du panel. Une erreur dans le logiciel a malheureusement interdit la sélection des enfants de 12 ans au cours des deux premiers trimestres. Pour compenser cette erreur, les enfants de douze ans ont bénéficié d'une probabilité de sélection égale au double de la précédente durant les trimestres trois et quatre. Le facteur de correction correspondait donc à la probabilité inverse, pour un membre quelconque du ménage, d'être sélectionné afin de faire partie du panel.

Correction pour les enfants de douze ans

Afin que les enfants de 12 ans soient correctement représentés, on a dû relever le facteur de pondération pour tenir compte des ménages où les enfants de cet âge n'avaient pu être sélectionnés, consécutivement au problème de logiciel. La correction correspond à la probabilité inverse, pour les enfants de 12 ans, d'être sélectionné dans les logements où une personne de 12 ans et plus devait faire partie du panel de l'étude longitudinale.

On se rappellera qu'à Montréal et dans les capitales régionales, les grappes ne sont couvertes qu'au cours d'un trimestre. Or, les enfants de 12 ans ne pouvaient être retenus durant les trimestres 1 et 2. Si on veut en tenir compte, la correction doit s'effectuer au niveau de la strate de l'ENSP et non de la grappe. Afin d'accroître la cohérence, pour les enfants de 12 ans, l'intégration et la correction se font au niveau de la strate de l'ENSP, peu importe la classe de l'ESS.

Correction pour la non-réponse du membre du panel

Cette correction tient compte des membres du panel des ménages répondants (c'est-à-dire des ménages qui ont répondu au questionnaire démographique) qui refusent de répondre au questionnaire qui leur est destiné. Elle prend la forme suivante :

GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION DE L'ENSP

somme des facteurs de pondération des répondants et des non-répondants du groupe âge-sexe

somme des facteurs de pondération des répondants du groupe âge-sexe

Le facteur de pondération de la formule correspond au facteur de pondération démographique multiplié par les corrections antérieures relatives aux membres du panel. Les groupes d'âge sont 12-24, 25-44, 45-64, 65 et plus pour les personnes des deux sexes. La correction s'effectue à l'échelon provincial.

Correction pour l'analyse comparative des membres du panel

Cette correction se calcule de la même façon que la correction pour l'analyse comparative des membres du ménage :

projections démographiques du groupe âge-sexe

somme des facteurs de pondération des répondants du groupe âge-sexe.

Les facteurs de pondération sont les facteurs de pondération démographiques, multipliés par les corrections apportées antérieurement dans la présente partie, y compris les facteurs de correction pour la non-réponse du membre du panel. Les groupes d'âge sont 12-24, 25-44, 45-64, 65 et plus, pour les sujets de sexe masculin ou féminin. Encore une fois, on a corrigé les valeurs obtenues pour tenir compte des régions boréales non couvertes par l'enquête.

On calcule le «**facteur de pondération du membre du panel**» en multipliant le facteur de pondération démographique par les facteurs de correction mentionnés dans la présente partie.