TABLE DES MATIÈRES

Cliquez sur le titre pour accéder au document Page

1. Introduction 7

2. Contexte 8

3. Objectifs 8

4. Contenu de l'enquête 9
   4.1 Critères de sélection 9
   4.2 Modifications apportées au contenu pour le cycle de 1996-1997 11
   4.3 Rétroaction des données de 1994-1995 et questions de suivi 13
   4.4 Nouveau contenu de l'enquête de 1996-1997 14

5. Plan d'échantillonnage 17
   5.1 Plan d'échantillonnage pour 1996-1997 17
      5.1.1 Plan d'échantillonnage de la composante principale des ménages 17
      5.1.2 Plan d'échantillonnage de 1996-1997 pour les suppléments provinciaux de la composante des ménages 21
      5.1.3 Aperçu de l'échantillonnage au moyen de la composition aléatoire 21
      5.1.4 Répartition de la taille de l'échantillon selon le secteur sanitaire 22
      5.1.5 Stratification et répartition de l'échantillon 24
   5.2 Plan d'échantillonnage pour 1994-1995 24
      5.2.1 Plan d'échantillonnage de la composante des ménages 25
      5.2.2 Répartition de l'échantillon 25
      5.2.3 Méthode d'exclusion 26
      5.2.4 Prélèvement de l'échantillon 26
      5.2.5 Plan d'échantillonnage au Québec 28

6. Collecte des données 31
   6.1 Conception du questionnaire et méthode de collecte des données 31
   6.2 Essais 31
   6.3 Interviews 32
   6.4 Surveillance et contrôle 33
   6.5 Non-réponse à l'ENSP 33
   6.6 Suivi des cas de non-réponse 33
   6.7 Dépistage 34

7. Traitement des données 35
   7.1 Vérification 35
   7.2 Codage 35

7.3 Création de variables dérivées et groupées 35
7.4 Facteurs de pondération 36
7.5 Suppression de renseignements confidentiels 36

8. Qualité des données 37
8.1 Taux de réponse 37
8.1.1 Taux de réponse transversaux - Échantillon principal 38
8.1.2 Taux de réponse transversaux - Échantillons sélectionnés par CA (par province et total) 39
8.1.3 Taux de réponse transversaux globaux 41
8.2 Erreurs relatives à l'enquête 42

9. Lignes directrices pour la totalisation, l'analyse et la diffusion 45
9.1 Lignes directrices pour l'arrondissement 45
9.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation 46
9.2.1 Définitions des genres d'estimations : de type nominal par opposition à quantitatives 46
9.2.2 Totalisation d'estimations de type nominal 48
9.2.3 Totalisation d'estimations quantitatives 48
9.3 Lignes directrices pour l'analyse statistique 49
9.4 Lignes directrices pour la diffusion 49

10. Tableaux de la variabilité d'échantillonnage approximative 51
10.1 Comment utiliser les tableaux de c.v. pour les estimations de type nominal 57
10.2 Exemples d'utilisation des tableaux de c.v. pour des estimations de type nominal 60
10.3 Comment utiliser les tableaux de c.v. pour calculer les limites de confiance 63
10.4 Exemple d'utilisation de tableaux de c.v. pour obtenir des limites de confiance 65
10.5 Comment utiliser les tableaux de c.v. pour effectuer un test Z 65
10.6 Exemple d'utilisation des tableaux de c.v. pour effectuer un test Z 66
10.7 Variances ou coefficients de variation réels 66
10.8 Seuils pour la diffusion des estimations relatives à l'ENSP 68

11. Pondération 75
11.1 Pondération de l'échantillon transversal de l'ENSP de 1996-1997 — Échantillon principal des ménages 76
11.1.1 Facteurs de pondération épurés 76
11.1.2 Corrections des facteurs de pondération appliqués aux membres des ménages 76
11.1.3 Corrections de la pondération pour les membres du panel 81

11.2 Pondération transversale de l'échantillon de l'ENSP de 1996-1997 — Provinces qui ont acheté un échantillon supplémentaire sélectionné par CA . . 83
11.2.1 Facteurs de pondération de base des unités sélectionnées par CA 85
11.2.2 Corrections supplémentaires des facteurs de pondération de base 86
11.2.3 Corrections supplémentaires des facteurs de pondération pour les membres du ménage 88
11.2.4 Corrections supplémentaires des facteurs de pondération pour les membres du panel 91
11.3 Méthodes de pondération fondées sur celles de 1994-1995 pour les provinces à part le Québec 95
11.3.1 Facteurs de pondération de base de l'EPA 96
11.3.2 Corrections supplémentaires des facteurs de pondération de base 96
11.3.3 Corrections supplémentaires pour les membres du panel 99
11.4 Méthodes de pondération pour le Québec fondées sur celles de 1994-1995 102
11.4.1 Facteurs de pondération de l'ESS 102
11.4.2 Facteur de pondération de base appliqué aux logements de l'ENSP 102
11.4.3 Corrections supplémentaires des facteurs de pondération de base 104
11.4.4 Corrections supplémentaires pour les membres du panel 105

12. Utilisation du fichier 109
12.1 Utilisation des facteurs de pondération 109
12.1.1 Facteur de pondération transversal - Fichier général - WT56 109
12.1.2 Facteur de pondération transversal - Fichier sur la santé - WT66 109
12.1.3 Facteur de pondération transversal - Fichier sur la santé - WT66_N 110
12.2 Convention appliquée pour nommer les variables 111
12.2.1 Structure élémentaire des noms des variables 111
12.2.2 Positions 1 et 2 : Nom de la variable / section du questionnaire 112
12.2.3 Position 3 : Type d'enquête 113
12.2.4 Position 4 : Variable de l'année de référence / du cycle 114
12.2.5 Position 5 : Type de variable 114
12.2.6 Positions 6 à 8 : Nom de la variable 115
12.3 Téléaccès aux fichiers principaux 115

Annexe A: Questionnaire

Annexe B: Cliché d'enregistrement - Fichier de microdonnées général

Annexe C: Cliché d'enregistrement - Fichier de microdonnées santé

Annexe D: Table alphabétique du dictionnaire des données - Fichier général
Index des variables du dictionnaire des données par sujet - Fichier général
Dictionnaire des données - Fichier de microdonnées général

Annexe E: Table alphabétique - Fichier de microdonnées santé
Index des variables par sujet - Fichier de microdonnées santé
Dictionnaire des données - Fichier de microdonnées santé

Annexe F: Variables dérivées et groupées

Annexe G: Tableaux de la variabilité, fichier de microdonnées général

Tableaux de la variabilité pour le Canada selon le groupe d'âge, fichier de microdonnées général

Tableaux de la variabilité selon les provinces et le Canada, fichier de microdonnées général

Tableaux de la variabilité selon le secteur sanitaire pour l'Ontario, le Manitoba et l'Alberta, fichier de microdonnées général

Annexe H: Tableaux de la variabilité, fichier de microdonnées santé

Tableaux de la variabilité pour le Canada selon le groupe d'âge, fichier de microdonnées santé

Tableaux de la variabilité selon les provinces et le Canada, fichier de microdonnées santé

Tableaux de la variabilité selon le secteur sanitaire pour l'Ontario, le Manitoba et l'Alberta, fichier de microdonnées santé

1. Introduction

Le programme de 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 a débuté en 1994; par la suite, d'autres collectes de données auront lieu tous les deux ans. Le programme servira à recueillir non seulement des données transversales, mais aussi des données longitudinales auprès d'un échantillon permanent de répondants interrogés tous les deux ans. Il comprend trois composantes : l'enquête auprès des ménages, l'enquête auprès des établissements et l'enquête dans le Nord.

La composante des ménages vise 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. La composante des établissements vise les résidents à long terme (plus de six mois) d'établissements de soins comportant quatre lits ou plus dans toutes les provinces, à l'exclusion du Yukon et des Territoires du Nord-Ouest. La composante du Nord vise les résidents des ménages du Yukon et des Territoires du Nord-Ouest, à l'exclusion de la population des réserves indiennes, des bases des Forces armées canadiennes et de certaines des régions les plus au nord des territoires.

Deux cycles de diffusion, à savoir 1994-1995 et 1996-1997, sont terminés pour la composante des ménages du programme de l'Enquête nationale sur la santé de la population. Le présent document a pour but de faciliter la manipulation des deux fichiers de microdonnées transversales pour 1996-1997, qui comprennent les résultats de l'enquête. Ces fichiers sont décrits de façon plus détaillée au chapitre 4 et dans les annexes.

Pour toute question au sujet des ensembles de données ou de leur utilisation, s'adresser à :

Pour un soutien technique ou des renseignements généraux sur les données :
Service d'aide aux utilisateurs de produits électroniques 1-800-949-9491

Pour des totalisations spéciales ou des renseignements généraux sur les données :
Services personnalisés à la clientèle de la Division des statistiques sur la santé 1-613-951-1746

Pour des renseignements sur le téléaccès : 1-613-951-1653
Sylvie Alary, ENSP Internet : nphs@statcan.ca

Pour des renseignements sur le contenu de l'enquête : 1-613-951-3285
Bryan Lafrance, ENSP Télécopieur : 1-613-951-4198

2. Contexte

À l'automne 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. Or, les sources existantes de données sur la santé ne fournissaient pas suffisamment de renseignements 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 les fonds nécessaires 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

Les objectifs décrits au chapitre 3 représentent seulement l'orientation générale de l'ENSP, notamment en ce qui concerne le type de renseignements à recueillir. À la première section, on expose les critères généraux utilisés pour choisir le contenu de l'enquête et on résume dans les grandes lignes le contenu des diverses sections et les changements. À la section suivante, on décrit en détail les modifications apportées au contenu de l'enquête pour réaliser le cycle de 1996-1997. La section qui suit met l'accent sur les variables de 1994-1995 qui ont fait l'objet d'une rétroaction et ont été utilisées pour le deuxième cycle. Enfin, à la dernière section, on décrit en détail le nouveau contenu pour l'enquête de 1996-1997.

4.1 Critères de sélection

On a choisi le contenu de l'enquête conformément aux critères suivants :

1) Les renseignements doivent être en rapport avec les objectifs des provinces et des territoires en matière de santé et faciliter le contrôle de ces objectifs. Dans les cas où ces objectifs n'existent pas, par exemple au niveau national, on peut déterminer le contenu de l'enquête à partir des politiques et des programmes.

2) Les renseignements ne doivent pas faire double emploi avec ceux provenant d'autres sources.

3) Pour faire mieux comprendre la santé et ses 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 porter principalement sur les comportements ou sur les problèmes de santé qui se prêtent à la prévention, au traitement ou à l'intervention.

5) L'enquête doit permettre de recueillir des renseignements sur les problèmes de santé 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 fournir des renseignements sur les facteurs liés non seulement à la maladie, mais aussi à la bonne santé.

Un certain nombre de renseignements ont été recueillis auprès de tous les membres de chaque ménage (composante générale). De plus, une personne a été choisie au hasard dans chaque ménage pour une interview plus approfondie (composante santé). Selon les critères susmentionnés, le questionnaire portait 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 a été mesuré à l'aide de questions sur la perception personnelle de la santé, de la capacité fonctionnelle, des problèmes de santé chroniques et de la limitation des activités. L'utilisation des services de santé a été mesurée au moyen de questions sur les visites faites à des professionnels de la santé, les soins hospitaliers et la consommation de médicaments. Les questions sur les facteurs de risque d'ordre comportemental avait trait à l'usage du tabac, à la consommation d'alcool et à l'activité physique. En outre, dans le cadre du deuxième cycle de l'enquête, on a accordé une attention spéciale à l'accès aux services. On a posé des questions sur les tests et les examens préventifs, ainsi que de questions d'approfondissement sur la fréquence de ceux-ci, les raisons justifiant leur utilisation ou leur non-utilisation et les obstacles rencontrés. Les questions sur les caractéristiques démographiques et socio-économiques portaient sur l'âge, le sexe, le niveau de scolarité, l'appartenance ethnique, le revenu du ménage et l'activité sur le marché du travail.

Afin de réduire le coût de la collecte des données, on a intégré l'Enquête sur la promotion de la santé (EPS) de 1996-1997 à l'ENSP au lieu de l'effectuer séparément à titre d'enquête complémentaire comme en 1994-1995. Les questions de cette enquête ont été posées à l'échantillon principal et à la majeure partie de l'échantillon choisi par composition aléatoire (CA). (Voir le chapitre 5, Plan d'échantillonnage, pour la description de l'échantillon principal et de l'échantillon choisi par CA.) Des questions supplémentaires détaillées portaient sur les sujets suivants : mesures prises pour améliorer la santé, poids souhaité, allaitement maternel, nutrition, VIH/SIDA, usage du tabac pendant la grossesse, opinions sur l'usage du tabac, opinions sur la consommation d'alcool, soutien social, santé sexuelle et sécurité routière.

Dans le cadre de l'entente sur l'achat d'unités d'échantillonnage supplémentaires conclue avec le ministère de la Santé de l'Alberta, des questions supplémentaires ont été posées à tous les répondants de l'Alberta (tant de l'échantillon principal que de l'échantillon choisi par CA) sur des sujets comme l'exposition au soleil, les sources d'information sur la santé, la violence et la sécurité personnelle. En outre, certaines questions qui faisaient partie de l'Enquête sur la promotion de la santé n'ont pas été posées à l'échantillon choisi par CA de l'Alberta, mais l'ont été aux répondants de l'échantillon principal.

Tant au Manitoba qu'en Alberta, dans le cadre de l'entente sur l'achat d'unités d'échantillonnage supplémentaires, lorsqu'un ménage comprenait un ou plusieurs enfants, un d'entre eux a été sélectionné pour faire partie de l'échantillon choisi par CA, en plus de l'adulte déjà sélectionné. Des questions sur les services de santé destinés aux enfants ont été posées à leur sujet.

L'annexe A contient la liste des questions de l'enquête.

4.2 Modifications apportées au contenu pour le cycle de 1996-1997

Les données de l'ENSP de 1996-1997 ont été recueillies principalement par téléphone, alors que celles du cycle précédent l'avaient été principalement par interview sur place. Par conséquent, on s'est efforcé de rendre les questions plus faciles à lire (pour les intervieweurs) et plus faciles à comprendre (pour les répondants) en utilisant un langage plus familier dans la mesure du possible. Sans remanier les sections, on a modifié les écrans, raccourci les catégories de réponse et créé des catégories précises pour les réponses survenant fréquemment dans la catégorie «Autre, précisez».

Composante Générale - Modifications par section

Utilisation des services de santé

UTIL-Q3 : les réponses mentionnées fréquemment dans la catégorie «Autre, précisez» en 1994-1995 ont été ajoutées à titre de nouvelles catégories.

UTIL-Q4 : la catégorie des groupes d'entraide a été supprimée de la liste des médecines non traditionnelles et a fait l'objet d'une question distincte, UTIL-Q4A.

UTIL-Q6 : le concept a été modifié. La question a été perfectionnée pour obtenir l'opinion de la personne pour laquelle on obtient les réponses au questionnaire par procuration et celle de la personne qui répond par procuration si la personne concernée a moins de 18 ans.

UTIL-Q7 et UTIL-Q10 ont été converties à des questions avec énumération fondées sur les réponses mentionnées le plus fréquemment lors de l'interview de 1994-1995.

Limitation des activités

RESTR-Q3 et RESTR-Q4 ont été regroupés en une seule question, avec un texte de remplissage donnant le libellé approprié fourni par l'ordinateur.

RESTR-Q6 a été éclaircie et on est passé d'une question où il fallait mentionner toutes les réponses pertinentes à une question où chaque forme d'aide est mentionnée séparément.

Problèmes de santé chroniques

CHR-Q1, qui était une question où on demandait d'indiquer toutes les réponses pertinentes, a été transformée en une série de questions individuelles pour chaque problème de santé. On a ajouté à la liste les troubles intestinaux et les troubles thyroïdiens. On a supprimé les sous-questions sur le type de cancer et sur la respiration sifflante de l'asthme.

Renseignements socio-démographiques

SOCIO-Q1 : on a ajouté la Hollande à la catégorie des Pays-Bas.

SOCIO-Q3 : on a modifié le concept légèrement en demandant quand le répondant est arrivé pour la première fois au Canada pour rester au lieu de demander l'année de l'immigration.

SOCIO-Q7 : on a modifié l'ordre des catégories de races et on les a perfectionnées de façon à refléter celles du Recensement de 1996.

Niveau de scolarité

EDUC-Q5 et EDUC-Q6 de 1994 1995 ont été placées en tête de la section pour devenir EDUC-Q1 et EDUC-Q2 afin de permettre la rétroaction des données précédentes. Par conséquent, EDUC-Q1 à EDUC-Q4 sont devenues EDUC-Q4 à EDUC-Q7, respectivement.

Population active

Cette section a été remaniée pour tirer parti des capacités du logiciel d'interview assistée par ordinateur (IAO) et de certains résultats de l'enquête de 1994-1995. On n'a posé les questions de cette section qu'aux personnes de 15 à 75 ans. On a réduit de six à trois le nombre d'emplois pour lesquels on a recueilli des données. On a éliminé la question LFS-Q1 sur l'activité principale, préférant laisser à l'application le soin de décider quel est l'emploi principal. On a redéfini intervalle de travail pour la faire passer de sept jours à 28 jours. Les questions LFS-Q17A et LFS-Q17B comptent quatre nouvelles catégories : retraite, démission, recherche d'un emploi et état d'invalidité/convalescence.

Composante Santé - Modifications par section

On a réorganisé les sections afin de mieux présenter les questions, principalement à cause de l'intégration de l'Enquête sur la promotion de la santé (EPS) et de l'échantillon supplémentaire de l'Alberta.

Taille et poids

Pour les adultes, les poids minimal et maximal permis pour l'IAO sont passés de 1 et 999 à une intervalle de 18 et 575, compte tenu des résultats du cycle de 1994-1995.

Accès aux services de santé

On a intégré les questions sur la prise de la tension artérielle, la mammographie et le test de Papanicolaou à cette section.

Activités physiques

Dans PA-Q1, on a précisé la catégorie «patinage» qui est devenue «patinage sur glace». Le patinage en patins à roues alignées ou le patinage à roulettes est maintenant inclus dans la catégorie «Autre, précisez». On a remplacé la catégorie «tai chi» par la catégorie «basketball».

Blessures

IN-Q7 n'a été posé que si le répondant travaillait au moment de l'interview.

Consommation de médicaments

DRG-Q1, qui était une question où il fallait indiquer toutes les réponses pertinentes, a été modifié pour devenir une série de questions individuelles pour chaque catégorie de médicament. L'ordre des catégories sur la prise d'hormones et sur la pilule anticonceptionnelle a été inversé. On a ajouté une nouvelle catégorie pour les médicaments destinés au traitement des troubles thyroïdiens.

Consommation d'alcool

ALC-Q2 : l'ordre des catégories de réponse a été inversé.

ALC-Q3 : les catégories de réponses ont été converties en une liste de fréquence (p. ex., une fois par mois, au lieu de demander un nombre précis).

Soutien social

SUP-Q7 : la catégorie «n'en a aucun» a été modifiée pour devenir «n'en avez pas ou habitent tous avec vous»

Numéro provincial d'assurance-maladie et administration

On a demandé au répondant longitudinal si son numéro d'assurance-maladie avait changé depuis la dernière interview. Le cas échéant, on a pris note du nouveau numéro. Autrement on a sauté la question.

4.3 Rétroaction des données de 1994-1995 et questions de suivi

Afin de réduire le fardeau de réponse, on n'a pas posé de nouveau les questions dont on connaissait déjà la réponse parce que celle-ci ne change pas au cours du temps (p. ex., pays de naissance). Pour les variables pour lesquelles la réponse peut changer au cours du temps, mais uniquement dans certaines conditions (p. ex., niveau de scolarité), la mise à jour n'a eu lieu qu'au besoin.

Limitation des activités

On s'est servi des données indiquant si le répondant souffrait ou non d'une incapacité en 1994. En cas de changement de situation, on a posé des questions pour obtenir une explication.

Problèmes de santé chroniques

À chaque répondant, on a rappelé les réponses concernant certains problèmes de santé chroniques (asthme, arthrite, hypertension, migraine, diabète, épilepsie, ulcères de l'estomac ou de l'intestin et les séquelles d'un accident vasculaire cérébral) pour essayer d'obtenir des explications concernant le changement. Si le problème était nouveau, on a enregistré la date de sa manifestation.

Renseignements socio-démographiques

Pour chaque répondant, on a entré de nouveau les indicateurs montrant qu'on a recueilli les renseignements concernant le pays de naissance et l'origine ethnique. Puisqu'on a modifié les catégories de réponse pour la race, on a de nouveau recueilli les renseignements pour cette variable. Enfin, on a de nouveau posé la question sur la première langue apprise et encore parlée, parce que la réponse à cette question peut changer avec le temps.

Niveau de scolarité

Pour chaque répondant, on a de nouveau entré un indicateur montrant le plus haut niveau de scolarité atteint. On s'est servi de questions filtres pour déterminer si le répondant avait fréquenté un établissement d'enseignement entre les deux cycles. Le cas échéant, on a de nouveau recueilli les renseignements sur le niveau de scolarité.

Population active

À chaque répondant, on a rappelé le nom de l'employeur, la branche d'activité et le genre de tâches correspondant à l'emploi principal mentionné à l'interview de 1994-1995. Aux répondants qui ont dit avoir travaillé l'année précédente, on a demandé de confirmer le nom de l'employeur.

Usage du tabac

Aux répondants longitudinaux, on a rappelé les renseignements donnés antérieurement concernant la consommation de cigarettes. Si l'interview de 1994-1995 n'avait pas eu lieu par procuration, on a posé des questions pour connaître les raisons du changement.

4.4 Nouveau contenu de l'enquête de 1996-1997

Composante Générale

Utilisation des soins de santé

On a inclus deux nouvelles questions sur les services de santé obtenus aux États-Unis.

Problèmes de santé chroniques

Pour certains problèmes de santé chronique (asthme, arthrite/rhumatisme, hypertension, migraine et diabète) on a posé des questions de suivi pour savoir si le répondant prenait un médicament ou suivait un traitement et, le cas échéant, pour obtenir des renseignements sur le type de traitement ou de médicament.

Composante Santé

Accès aux services de santé

En 1996-1997, on a mis principalement l'accent sur l'accès aux services de santé. Pour divers services de santé, on a posé des questions sur la fréquence d'utilisation, les raisons de l'utilisation, les obstacles rencontrés et les raisons pour lesquelles les services ne sont pas utilisés ou sont utilisés moins fréquemment que ne le recommande l'Association médicale canadienne. Ces services incluent la mesure de la tension artérielle, le test de Papanicolaou, la mammographie, l'examen médical, la vaccination antigrippale, les visites chez le dentiste et l'examen de la vue.

Mouvements répétitifs

Les nouvelles questions sur les mouvements répétitifs ont été posées juste avant la section sur les blessures, afin d'isoler les blessures consécutives à ce genre de mouvements.

Consommation de médicaments

Aux répondantes qui ont dit suivre une hormonothérapie, on a demandé quel type d'hormones elles prenaient et quelle année a débuté le traitement.

Dépendance à l'égard de l'alcool

Ces questions font partie de la série élaborée par Kessler et Mroczek pour l'échelle abrégée de sélection pour 12 mois du *Composite International Diagnostic Interview (CIDI)*.

Santé mentale

On a ajouté une question de suivi que l'on a posée aux répondants qui ont dit avoir consulté un médecin au cours des 12 derniers mois. Il s'agit d'une liste de personnes que le répondant pourrait avoir consultées ou à qui il pourrait avoir parlé.

Administration

Les renseignements de contact concernant le répondant longitudinal ont été recueillis dans le cadre de la composante santé. Le numéro de téléphone au travail est une nouvelle variable.

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.

5.1 Plan d'échantillonnage pour 1996-1997

5.1.1 Plan d'échantillonnage de la composante principale des ménages

Au cours du premier cycle de l'ENPS, l'échantillon a été établi en choisissant tout d'abord des ménages, puis à l'intérieur de chacun d'eux, un membre devant devenir le répondant longitudinal. Pour le deuxième cycle, une distinction est faite entre l'échantillon choisi à des fins longitudinales et celui choisi à des fins transversales.

L'échantillon longitudinal de 1996-1997 comprend tous les répondants longitudinaux choisis au cours du premier cycle qui ont répondu au moins à la composante générale du questionnaire en 1994-1995. Cet échantillon comprend 2 022 personnes de moins de 12 ans au moment du premier cycle [qui avaient été interviewées dans le cadre de l'Enquête nationale longitudinale sur les enfants et les jeunes (ELNEJ)] qui ont été incluses dans l'échantillon de l'ENSP de 1996-1997. Les unités sélectionnées en 1994-1995, dans le cadre de l'achat d'unités d'échantillonnage supplémentaires, ont été exclues. On a uniquement effectué le dépistage des répondants longitudinaux, à partir des renseignements de base recueillis en 1994-1995; aucune tentative n'a été faite pour dépister tous les membres du ménage au fil des ans. Aux fins de la production des données transversales, tous les membres du ménage vivant avec le répondant longitudinal ont été interviewés. Les données relatives au répondant longitudinal choisi seront utilisées à la fois à des fins longitudinales et transversales.

L'échantillon principal choisi en 1994-1995 n'a pas été élargi en vue de la production des estimations transversales pour 1996-1997. Toutefois, dans trois provinces, des unités d'échantillonnage additionnelles ont été achetées à des fins longitudinales (voir 5.1.2). Dans les provinces où il n'y a pas d'unités supplémentaires, la taille de l'échantillon est légèrement inférieure à celle de 1994-1995, en raison de l'attrition, des décès, des cas hors du champ d'observation, des cas impossibles à retracer et des non-répondants. Cette diminution de la taille de l'échantillon, relativement faible, ne devrait pas augmenter considérablement la variance des estimations. L'analyse de l'échantillon de 1996-1997 avant la collecte des données a montré que sa représentativité selon la province, l'âge ou le sexe n'était pas biaisée.

Taille de l'échantillon longitudinal principal selon la province courante
Nombre de personnes

Province	Échantillon principal de 1994-1995[1]	Échantillon principal de 1996-1997	% de suivi de l'échantillon principal de 1994-1995
Terre-Neuve	1 212	1 082	89,3
Île-du-Prince-Édouard	1 184	1 037	87,6
Nouvelle-Écosse	1 271	1 085	85,4
Nouveau-Brunswick	1 277	1 125	88,1
Québec	3 430	3 000	87,5
Ontario	5 335	4 307	80,7
Manitoba	1 346	1 205	89,5
Saskatchewan	1 320	1 168	88,5
Alberta	1 697	1 544	91,0
Colombie-Britannique	2 023	1 723	85,2
Total	20 095	17 276	86,0

Comme aucune nouvelle unité longitudinale n'a été sélectionnée au cours du deuxième cycle, la population comprise dans l'échantillon longitudinal en 1996-1997 est âgée de deux ans ou plus. Cela veut dire qu'aucune personne sélectionnée n'est âgée de 0 ou de 1 an au cours du deuxième cycle et, par conséquent, qu'aucune estimation directe des variables comprises dans la composante santé du questionnaire pour l'échantillon principal ne sera possible pour ces âges. La non-sélection de nouvelles unités longitudinales au cours du deuxième cycle a aussi eu pour effet que les personnes qui ont immigré au Canada au cours des deux dernières années (depuis la dernière sélection des unités longitudinales dans le premier cycle) ne sont pas représentées par les personnes choisies pour faire partie de l'échantillon longitudinal. Toutefois, des estimations des variables de la composante générale sont disponibles pour l'ensemble de la population transversale de 1996-1997, y compris le groupe des 0 et 1 an. Les immigrants récents qui se sont joints à des ménages ayant participé à l'enquête en 1994-1995 sont aussi représentés dans l'échantillon principal. Il convient de souligner que la stratification a posteriori en fonction des totaux de la population de 1996-1997 (voir le chapitre 11 — Pondération) traite de façon implicite la population immigrante non couverte, comme si elle avait les mêmes caractéristiques que le reste de la population.

[1] Échantillon réel, en excluant les logements non admissibles, vacants, en construction ou rejetés par suite de l'application de la méthode d'exclusion.

Taille de l'échantillon transversal principal selon la province courante (à l'exclusion des suppléments provinciaux)

Nombre de personnes

Province	Composante Générale	Composante Santé
Terre-Neuve	3 017	1 082
Île-du-Prince-Édouard	2 752	1 037
Nouvelle-Écosse	2 775	1 085
Nouveau-Brunswick	2 888	1 125
Québec	7 838	3 000
Ontario	10 899	4 307
Manitoba	3 045	1 205
Saskatchewan	2 785	1 168
Alberta	4 164	1 544
Colombie-Britannique	4 276	1 723
Total	44 439	17 276

En 1994-1995, la population cible comprenait les résidents des ménages de toutes les provinces, à l'exclusion des personnes vivant dans les réserves indiennes, les bases des Forces armées canadiennes et les régions éloignées de l'Ontario et du Québec. Au niveau transversal, la population cible principale comprend tous les membres des ménages qui vivent avec des répondants longitudinaux. La population cible des suppléments provinciaux est décrite à la section 5.1.2.

Pour déterminer si une personne fait ou non partie de la population cible, on a établi les règles suivantes aux fins de la collecte des données.

Situation du répondant longitudinal	**Mesures prises**
Décédé	Dans le cas des répondants longitudinaux dont on a déterminé qu'ils étaient décédés au moment de l'interview du deuxième cycle, le décès a été confirmé à partir de la Base canadienne de données sur la mortalité. Les répondants longitudinaux décédés font partie du fichier longitudinal, mais ils ne figurent pas dans les fichiers transversaux.
Placé dans un établissement	Les répondants longitudinaux qui sont passés d'un ménage privé à un établissement de soins de santé ont été interviewés dans le cadre de la composante des établissements de l'ENSP. Les données les concernant figurent dans le fichier longitudinal des ménages, mais pas dans le fichier transversal.
Déménagé dans les Territoires du Nord-Ouest ou au Yukon	Dans les cas où le répondant longitudinal avait déménagé, on a tenté de trouver son nouveau domicile. Dans la mesure du possible, le répondant longitudinal a été interviewé au moyen du questionnaire de base des ménages. Ces répondants font partie du fichier longitudinal, mais pas des fichiers transversaux.
Déménagé dans une réserve indienne ou dans une base des Forces armées canadiennes	Si, après dépistage, on a retrouvé le répondant longitudinal dans une réserve indienne ou une base des Forces armées canadiennes, on s'est efforcé d'interviewer la personne sélectionnée et les autre membres du ménage en se servant du questionnaire visant les ménages. Ces répondants font partie du ficher longitudinal, mais ne figurent pas dans le fichier transversal.
Déménagé à l'extérieur du Canada de façon temporaire	Dans les cas où le répondant longitudinal avait déménagé à l'extérieur du Canada, pour une période donnée, mais devait revenir y vivre, on a tenté de l'interviewer. Encore une fois, ces unités font partie du fichier longitudinal, mais pas du fichier transversal.
Déménagé à l'extérieur du Canada de façon permanente	Si le répondant longitudinal avait quitté le Canada et ne devait pas y revenir, on a recueilli des données au sujet de son nouveau lieu de résidence, et celles-ci pourraient être utilisées pour le suivi des vagues subséquentes. Ces personnes n'ont pas été interviewées et ne figurent ni dans le fichier longitudinal ni dans les fichiers transversaux.

5.1.2 Plan d'échantillonnage de 1996-1997 pour les suppléments provinciaux de la composante des ménages

Dans trois provinces, l'Alberta, l'Ontario et le Manitoba, les gouvernements provinciaux ont fourni des fonds additionnels, afin qu'un échantillon plus large des logements puisse être sélectionné. Cet achat d'unités supplémentaires avait pour but d'obtenir un échantillon de taille suffisante pour produire des estimations transversales fiables au niveau infraprovincial (secteurs sanitaires). L'échantillon supplémentaire est combiné à l'échantillon principal pour produire un grand fichier transversal des données.

Toutes les interviews ont été effectuées par téléphone. Un échantillon de numéros de téléphone a été sélectionné à partir du système de composition aléatoire (CA) de Statistique Canada. Autrement dit, les personnes qui n'ont pas le téléphone n'avaient aucune chance d'être sélectionnées. En outre, les personnes vivant dans les réserves indiennes, les bases des Forces armées canadiennes, les établissements de soins ou les logements collectifs n'étaient pas admissibles à l'interview. Les méthodes de pondération permettent le rajustement nécessaire pour tenir compte des différences de couverture de la population. On a défini la taille de l'échantillon selon le secteur sanitaire d'après le financement disponible et les exigences des provinces quant à l'obtention d'estimations fiables selon le secteur sanitaire.

Comme pour l'enquête principale, on a administré une composante générale à tous les membres des ménages répondants. Dans les trois provinces, un membre du ménage âgé de 12 ans ou plus a été sélectionné pour répondre à la composante santé. En Alberta et au Manitoba, un enfant de 0 à 11 ans a aussi été sélectionné (dans la mesure du possible) et a répondu au même questionnaire que les enfants choisis dans l'échantillon principal.

5.1.3 Aperçu de l'échantillonnage au moyen de la composition aléatoire

Le système d'échantillonnage au moyen de la composition aléatoire de Statistique Canada repose sur une méthode appelée élimination des banques non valides, qui fonctionne de la manière suivante. Une banque (l'indicatif régional plus les cinq premiers chiffres d'un numéro de téléphone) est considérée comme admissible aux fins de l'échantillonnage si elle comprend au moins un numéro de téléphone résidentiel, comme en fait foi un fichier des factures de téléphone résidentielles fourni par les compagnies de téléphone. Les banques admissibles sont regroupées pour former des strates. À l'intérieur d'une strate, on choisit une banque au hasard et un numéro entre 00 et 99 est produit de façon aléatoire. La combinaison de la banque et de ce numéro aléatoire produit un numéro de téléphone à sept chiffres. Un fichier historique permet de s'assurer que ce numéro n'a pas été utilisé récemment dans le cadre d'une autre enquête téléphonique à CA.

Le processus est repris jusqu'à ce que l'on ait produit le nombre de numéros de téléphone souhaités à l'intérieur de la strate. Souvent, le numéro généré correspond à un numéro de téléphone qui n'est pas en service. Cela signifie que pour atteindre les objectifs fixés quant à la taille de l'échantillon, de nombreux numéros de téléphone additionnels doivent être produits. Le taux de succès global se situe généralement autour de 50 %, ce qui signifie qu'environ la moitié des numéros générés correspondent réellement à des ménages. Évidemment, ces chiffres varient selon la région.

5.1.4 Répartition de la taille de l'échantillon selon le secteur sanitaire

La taille des échantillons a été déterminée de concert avec les ministères de la Santé concernés, à partir de deux critères : les fonds disponibles et la précision des estimations requises. Les secteurs sanitaires ont été définis par les ministères provinciaux de la Santé à partir des données géographiques du Recensement de 1991.

En Ontario, il a été décidé que pour chacun des 23 secteurs sanitaires on avait besoin d'un total de 1 200 répondants pour la composante santé, lesquels accepteraient que les données les concernant soient fournies au ministère de la Santé. Des exceptions à cette règle se sont produites dans les secteurs sanitaires comprenant Toronto et Ottawa. Dans ces régions, le nombre requis de répondants était de 3 000 et 2 000 respectivement.

Au Manitoba, le nombre de répondants choisis par CA a été établi de façon que le nombre de répondants de l'échantillon principal et le nombre de répondants de l'échantillon choisi par CA combinés permettent d'atteindre un certain seuil. Dans le secteur sanitaire comprenant Winnipeg, le total des deux composantes devait représenter au moins 1 428 répondants pour la composante santé. Dans les deux secteurs sanitaires du Nord, le total a été fixé à 603. Dans les huit autres secteurs, il a été établi à 1 000. Le nombre de répondants souhaités pour l'échantillon principal était d'environ 30 dans le Nord, de 30 à 110 dans les secteurs au sud, à l'exclusion de Winnipeg, et d'environ 325 pour le secteur sanitaire de Winnipeg. L'échantillon choisi par CA devait correspondre au reste des répondants.

En Alberta, on a calculé le nombre de répondants requis selon la méthode qui suit. Dans chacun des 17 secteurs sanitaires, on a défini un niveau de précision pour les estimations. La taille de l'échantillon devait être suffisante pour atteindre ce niveau de précision pour une caractéristique observée dans 15 % de la population au niveau démographique de le secteur infrasanitaire. On a choisi pour niveau démographique soit le sexe, soit le groupe d'âge (0 à 11 ans, 12 à 24 ans, 25 à 44 ans, 45 à 64 ans et 65 ans et plus), ou le groupe d'âge/sexe (un croisement des deux groupes précédents).

Par exemple, dans le secteur sanitaire de Calgary, il fallait qu'une caractéristique observée dans 15 % de la population ait un coefficient de variation (c.v.) ne dépassant pas 25 % des estimations au niveau de l'âge/du sexe. Dans les secteurs sanitaires du Nord, le c.v. a aussi été fixé à 25 %, mais ce niveau de précision ne devait être atteint que pour le sexe. Les effets du plan d'échantillonnage et le nombre prévu de répondants dans chaque groupe démographique du secteur sanitaire ont été utilisés pour déterminer le nombre de ménages requis, celui-ci allant de 230 à 2 450 par secteur sanitaire, selon les exigences.

Le tableau figurant ci-dessous montre les totaux provinciaux (nombres de répondants et numéros de téléphone approximatifs)

Taille des échantillons supplémentaires choisis par CA selon la province

Province	Nombre de secteurs sanitaires dans les fichiers de microdonnées	Nombre de secteurs sanitaires	Taux de réponse visé	Nombre prévu de répondants	Nombre de numéros de téléphone générés
Ontario	23/16	23	75 %	32 000	89 500
Manitoba	5	11	80 %	10 300	29 000
Alberta	5	17	80 %	13 000	30 300

5.1.5 Stratification et répartition de l'échantillon

Les circonscriptions (c.-à-d. les trois premiers chiffres du numéro de téléphone) ont été utilisées comme unités de stratification, et les strates ont été composées de groupes de circonscriptions. Grâce aux codes postaux figurant dans les fichiers de factures de téléphone, un emplacement géographique a été déterminé pour chaque numéro de téléphone.

Les strates primaires étaient constituées de circonscriptions portant sur les mêmes secteurs sanitaires. Une règle fondée sur le pourcentage de numéros compris dans la circonscription pour chaque secteur sanitaire a été établie, afin de déterminer combien de secteurs étaient couverts de façon significative par une circonscription. La plupart des circonscriptions ne couvraient qu'un secteur sanitaire, mais certaines en couvraient deux, et quelques-unes, trois. Un taux d'échantillonnage pour chaque secteur a été déterminé à partir du nombre requis de numéros de téléphone échantillonnés, du taux de succès et du nombre estimatif de numéros de téléphone compris dans le secteur sanitaire.

Dans les strates qui ne couvraient de façon significative qu'un secteur sanitaire, l'échantillonnage s'est fait selon le taux déterminé pour le secteur sanitaire concerné. Dans le cas des strates qui couvraient plus d'un secteur sanitaire de façon significative, on a utilisé le taux d'échantillonnage le plus élevé parmi les secteurs sanitaires couverts de façon significative. Les strates primaires devaient être suffisamment importantes pour qu'un échantillon d'au moins 40 ménages puisse être établi. Si la taille était inférieure, la strate a été regroupée avec une autre couvrant des secteurs sanitaires similaires. Dans les strates primaires plus importantes, on a procédé à une sous-stratification géographique. Les échantillons tirés des states finales devaient comprendre de 40 à 500 ménages environ.

5.2 Plan d'échantillonnage pour 1994-1995

L'Enquête sur la population active (EPA) remaniée a servi de base pour le plan d'échantillonnage dans toutes les provinces, à l'exception du Québec, où l'on a choisi un échantillon à partir des ménages déjà interviewés par Santé Québec pour l'Enquête sociale et de santé de 1992-1993.

5.2.1 Plan d'échantillonnage de la composante des ménages

Le plan d'échantillonnage de la composante des ménages a été établi selon trois 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'EPA remaniée comme véhicule pour sélectionner l'échantillon.

Ces 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.

5.2.2 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)[1]. Selon cette méthode, l'échantillon a été réparti proportionnellement à $\propto(0.804W_h^2 + 1/12^2)$, où W_h représente la proportion des ménages dans la province/le territoire (h, h=1,..,12) d'après le Recensement de 1991. Cette répartition a permis de déterminer la taille de l'échantillon principal de chaque province. Quatre provinces ont choisi d'acheter des unités d'échantillonnage additionnelles 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 additionnelles 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 additionnelles pour satisfaire à des exigences minimales dans leurs secteurs sanitaires, tandis que le Nouveau-Brunswick et la Colombie-Britannique en ont acheté pour certains secteurs seulement. En Colombie-Britannique, presque toutes les unités d'échantillonnage additionnelles 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 des échantillons a été élargie en fonction du nombre de ménages que l'on s'attendait de rejeter à la sélection.

5.2.3 Méthode d'exclusion

Le contenu de l'enquête vise principalement un membre de chaque ménage échantillonné 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, sans le recours à la méthode d'exclusion, les personnes appartenant à des ménages de grande taille, notamment les parents et leurs enfants, seraient 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, seraient surreprésentées.

Par conséquent, on a adopté une méthode d'exclusion pour accroître la représentativité des parents et des jeunes dans le panel. Une partie de l'échantillon a été 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 ont été considérés comme remplissant les conditions de rejet et exclus de l'enquête. Afin de maintenir la taille voulue des échantillons, on a ajouté au nombre de ménages visités dans chaque province le nombre prévu de ménages 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.

5.2.4 Prélèvement de l'échantillon

Un plan d'échantillonnage stratifié à deux degrés a été utilisé pour la composante des ménages de l'ENSP. Le premier degré consistait à former des strates homogènes et à prélever des échantillons indépendants de grappes dans chaque strate. Le deuxième degré consistait 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.

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 de l'EPA 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.

5.2.5 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 secteurs 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 secteur, 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.

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 $\%(2W_h^2 + 1/4^2)$, 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'ELNEJ 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 sont survenus dans certains de ces logements.

Bibliographie

[1] Kish, L. (1988). Multipurpose Sample Design, *Survey Methodology*, 14, 19-32.

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 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. 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 termes, 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 d'étudier divers aspects de son contenu. Les principaux objectifs de ces discussions étaient de vérifier la clarté et la qualité des questions, ainsi que les réactions des répondants aux sections jugées délicates (VIH, 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. Pour l'échantillon principal, les interviews ont été menées par des intervieweurs expérimentés de l'Enquête sur la population active. Pour l'échantillon choisi par CA, en raison de l'augmentation du volume, on n'a pas recouru à des intervieweurs de l'EPA dans tous les cas. Ces deux essais visaient encore une fois à observer les réactions des répondants, à estimer la durée des diverses sections du questionnaire et à déterminer les taux de réponse. Les opérations et les procédures sur le

terrain, la formation des intervieweurs et le programme d'ordinateur (c'est-à-dire 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 à grande échelle à l'interne pour repérer toute erreur dans le cheminement du programme ou dans le texte. 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

La collecte des données pour l'échantillon principal a été divisée en quatre trimestres (juin, août et novembre 1996 et févier 1997). Les intervieweurs sont des employés à temps partiel recrutés et formés expressément pour mener des enquêtes selon la méthode d'interview assistée par ordinateur. Une autre collecte s'est tenue en juin 1997, avec des tentatives supplémentaires de dépistage des non-répondants des trimestres précédents.

La collecte auprès des échantillons choisis par CA a été effectuée mensuellement, les dates de début et de fin étant fonction du financement fourni par la province concernée. Dans le cas de l'Alberta, la collecte s'est déroulée de juin 1996 à mars 1997. Pour l'Ontario, elle est allée d'octobre 1996 à août 1997. Au Manitoba, elle s'est déroulée de novembre 1996 à août 1997. Des interviews ont été aussi été effectuées par des employés à temps partiel, généralement des intervieweurs d'expérience travaillant par téléphone, qui ont été recrutés et formés expressément pour interviewer les échantillons choisis par CA au moyen de la méthode d'interview assistée par ordinateur.

On a d'abord communiqué par téléphone avec les répondants de l'échantillon principal, et 95 % des interviews ont été effectuées par téléphone. On a effectué des interviews en personne pour les répondants qui n'avaient pas le téléphone, dans le cadre du dépistage d'un répondant, ou à la demande d'un répondant. Pour la collecte auprès de l'échantillon choisi par CA, aucune interview en personne n'était permise. L'interview complète a duré en moyenne une heure auprès de chaque ménage.

Dans chaque logement, des renseignements sur tous les membres du ménage ont été recueillis auprès de la personne qui se trouvait à la maison au moment de l'appel de l'intervieweur. On a recours à de telles déclarations par «personne interposée», qui représentent environ 55 % des renseignements recueillis, pour réduire le coût des opérations de collecte de la composante générale.

Dans le cas des répondants choisis, les déclarations par personne interposée pour la composante santé ont été permises seulement pour des raisons de maladie ou d'incapacité. Ces déclarations par personne interposée représentent 2 % des données recueillies pour les

répondants de 12 ans et plus. En revanche, toutes les interviews auprès des répondants choisis âgés de moins de 12 ans ont été effectuées par personne interposée.

6.4 Surveillance et contrôle

Dans le cas de l'échantillon principal, on a appliqué la structure de surveillance et de contrôle de l'EPA à la collecte des données de l'ENSP. Pour l'échantillon sélectionné par CA, on a créé une structure similaire d'intervieweurs, d'intervieweurs principaux et de gestionnaires de projet. Chaque projet concernant un échantillon choisi par CA a été mené d'un bureau central, afin que les gestionnaires soient sur place pour assurer une supervision constante.

6.5 Non-réponse à l'ENSP

Les intervieweurs ont eu pour instructions de faire tous les efforts raisonnables pour interroger les membres des ménages admissibles à l'ENSP. Si une personne faisant partie de l'échantillon principal refusait dès le départ de participer à l'ENSP, le bureau régional envoyait une lettre au répondant longitudinal pour souligner l'importance de l'enquête et de la coopération du ménage. Ensuite, l'intervieweur faisait un deuxième appel (ou une deuxième visite). Lorsque le moment de l'appel (ou de la visite) de l'intervieweur était mal choisi, ce dernier fixait un rendez-vous pour rappeler ou revenir à un moment plus opportun. Lorsqu'il n'y avait personne à la maison, de nombreuses autres tentatives ont été faites pour contacter le ménage. Dans le cas de l'échantillon choisi par CA, plusieurs appels ont été faits aux numéros de téléphone non répondants. Au besoin, une lettre comprenant des renseignements au sujet de l'enquête a été envoyée par le bureau régional aux répondants, pour souligner l'importance de l'enquête et de la coopération du ménage. Les ménages non répondants compris dans l'échantillon n'ont jamais été remplacés par d'autres ménages.

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 des interviews, la taille de la tâche confiée à chaque intervieweur a été déterminée 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).

La formation offerte aux intervieweurs englobait des moyens de réduire le nombre de cas de non-contact (par exemple, faire des visites ou des appels à différentes heures de la journée) en se servant des renseignements de base recueillis au cours de l'interview précédente (pour l'échantillon principal) et de l'assistance-annuaire pour valider le numéro (pour l'échantillon choisi par CA).

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.

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

6.7 Dépistage

Dans le cas de l'échantillon principal, le fait de ne pas dépister un répondant longitudinal a été considéré comme un autre type de non-réponse. Les intervieweurs disposaient de plusieurs façons pour dépister un répondant. Les données relatives au cas comprenaient la dernière adresse et le dernier numéro de téléphone, ainsi que le nom et l'adresse d'un ou deux contacts, si ces renseignements avaient été recueillis. En outre, on a formé les intervieweurs pour qu'ils se servent des outils disponibles, par exemple, les annuaires téléphoniques locaux et l'assistance-annuaire. Lorsque ces pistes n'aboutissaient pas, on a transmis les cas à un intervieweur d'expérience ayant une formation particulière pour le dépistage des répondants. Les intervieweurs de dépistage avaient accès aux répertoires téléphoniques de l'ensemble du Canada et à des annuaires par numéros. Le taux de non-réponse découlant de l'impossibilité de dépister le répondant longitudinal se chiffre à 1,7 %, ce qui est extrêmement faible.

7. Traitement des données

7.1 Vérification

La vérification est exécutée en direct par la voie du programme d'interview assistée par ordinateur (IAO) durant la collecte des données. Il est impossible de saisir des valeurs hors champ et les erreurs d'enchaînement sont contrôlées à l'aide du système d'IAO. On ne peut entrer des valeurs incorrectes et l'enchaînement correct des questions se fait automatiquement. Par exemple, le système d'IAO veille à ce que toute question ne s'appliquant pas au répondant ne soit pas posée. 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 générale, 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.

7.2 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.

7.3 Création de variables dérivées et groupées

Pour simplifier l'analyse des données, on a calculé plusieurs variables du fichier de microdonnées en utilisant des items du questionnaire de l'ENSP. Le cinquième caractère du nom des variables dérivées est habituellement un D ou un G. Dans certains cas, les variables dérivées sont simplement obtenues en groupant des catégories de réponse. Dans d'autres, plusieurs variables sont combinées pour en former une nouvelle. Les détails du calcul de ces variables plus complexes figurent à l'annexe F.

7.4 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 SMC6_2 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.

7.5 Suppression de renseignements confidentiels

Il convient de mentionner que le fichier de microdonnées «à grande diffusion» 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. Ces mesures comptent l'élimination de valeurs individuelles, le groupement des variables et le plafonnement de la valeur des variables. Les utilisateurs qui désirent obtenir des renseignements exclus du fichier de microdonnées ont deux options : acheter des totalisations personnalisées ou avoir recours à l'option téléaccès qui permet aux utilisateurs de soumettre leur programme informatique pour qu'il soit traité à Statistique Canada. Pour plus d'information sur le téléaccès, voir le chapitre 12. Les estimations produites seront communiquées à l'utilisateur, à condition qu'elles soient conformes aux lignes directrices pour l'analyse et la diffusion décrites au chapitre 9 du présent document.

8. Qualité des données

8.1 Taux de réponse

Le calcul des taux de réponse comprend deux éléments uniques à l'ENSP. Le premier problème tient à l'acquisition, en Alberta, au Manitoba et en Ontario, d'unités d'échantillonnage supplémentaires sélectionnées par CA qui ont servi aux estimations transversales uniquement. On doit donc calculer séparément les taux de réponse transversaux pour l'échantillon principal, pour la partie sélectionnée par CA et pour l'échantillon global. Les taux de réponse calculés pour 1996-1997 sont résumés dans le tableau ci-après.

Tableau des taux de réponse pour 1996-1997

Niveau	Échantillon principal	Échantillon CA	Global (Principal +CA)
Ménages	94,3 %	80,0 %	82,6 %
Membres choisis : sauf les enfants choisis par CA	98,7 %	94,8 %	95,6 %
Membres choisis : enfants choisis par CA uniquement		98,2 %	98,2 %

Suit la description de la méthode de calcul du taux de réponse des ménages et du taux de réponse des membres des ménages sélectionnés pour faire partie du panel appliquée aux fichiers transversal et longitudinal. Il convient de souligner que les logements hors du champ de l'enquête (logements vacants ou abandonnés, logements en construction ou ménages ne remplissant pas les conditions requises pour faire partie de l'échantillon) n'ont été pris en compte dans aucun calcul. On notera aussi que, dans les trois sections qui suivent, on calcule le taux de réponse des membres du panel d'après le nombre de ménages répondants. Pour avoir une idée du taux global de réponse des membres du panel (fondé sur l'ensemble des ménages), il suffit de multiplier entre eux les deux taux.

8.1.1 Taux de réponse transversaux - Échantillon principal

***Taux de réponse des ménages* (M)**

$$\frac{\textit{nombre de ménages répondants}}{\textit{tous les ménages faisant partie du champ de l'enquête}}$$

On calcule le taux de réponse de 1996-1997 pour l'échantillon *principal transversal* en se fondant sur l'ensemble des ménages qui continuent à participer à l'enquête, sauf ceux hors champ de l'enquête (p. ex., un membre du panel qui a déménagé à l'étranger). Un ménage répondant est un ménage dont *au moins* un membre a répondu à la composante générale du questionnaire (HO5). Le taux de réponse transversal à l'ENSP, au niveau national, se chiffre à ***94,3 %.*** Au niveau provincial, ce taux varie de 91,8 % en Colombie-Britannique à 96,5 % à Terre-Neuve.

***Taux de réponse des membres du panel* (MP)**

On peut se représenter le taux de réponse des membres du panel comme le nombre de composantes santé (HO6) qui ont *effectivement* été remplies comparativement au nombre qui *auraient dû* être remplies.

$$\frac{\textit{nombre de personnes qui ont répondu au questionnaire H06}}{\textit{nombre de ménages répondants (c.-à-d. remplissant les conditions requises pour répondre)}}$$

Pour l'échantillon principal, le taux de réponse des membres du panel à l'ENSP se chiffre à ***98,7 %*** au niveau national et varie de 98,3 % au Québec à 99,3 % au Nouveau-Brunswick et en Alberta.

Renseignements pertinents pour le calcul des taux de réponse

Nombre de répondants au niveau des ménages :	16 215
Nombre de répondants au niveau du panel :	15 681
Nombre de non-répondants au niveau des ménages :	974
Nombre de non-répondants au niveau du panel :	206
Nombre de ménages hors du champ de l'enquête :	87

Calcul du taux de réponse des ménages :

$$\text{Taux M} = \frac{16\,215}{16\,215 + 974} = \frac{16\,215}{17\,189} = 94{,}3\ \%$$

Calcul du taux de réponse des membres du panel :

$$\text{Taux MP} = \frac{15\,681}{15\,681 + 206} = \frac{15\,681}{15\,887} = 98{,}7\ \%$$

8.1.2 Taux de réponse transversaux - Échantillons sélectionnés par CA (par province et total)

Taux de réponse des ménages

$$\frac{\textit{nombre de ménages répondants}}{\textit{tous les ménages faisant partie du champ de l'enquête}}$$

Pour tous les ménages des échantillons sélectionnés par CA uniquement, le taux de réponse pour les trois provinces regroupées est égal à 80,0 %. Au niveau provincial, les taux sont :

Ontario = 37 795 / 48 770 = 77,5 %
Manitoba = 11 210 / 12 952 = 86,6 %
Alberta = 13 531 / 16 459 = 82,2 %

Pour l'échantillon sélectionné par CA, les ménages hors champ de l'enquête incluent, par exemple, les numéros hors d'usage ou les numéros d'affaires.

Taux de réponse des membres du panel autres que les enfants (3 provinces)

$$\frac{\textit{nombre de répondants autres que les enfants au questionnaire H06}}{\textit{nombre de ménages répondants (c.-à-d. remplissant les conditions requises pour répondre)}}$$

Le taux de réponse des membres du panel qui ne sont pas des enfants pour les trois provinces regroupées est 94,8 %. Au niveau provincial, les taux sont :

Ontario = 35 527 / 37 795 = 94,0 %
Manitoba = 10 840 / 11 210 = 96,7 %
Alberta = 12 925 / 15 531 = 95,5 %

Taux de réponse des enfants membres du panel - (2 provinces uniquement)

$$\frac{\textit{nombre d'enfants qui ont répondu au questionnaire H06}}{\textit{nombre de ménages répondants comptant un enfant de moins de 12 ans (c.-à-d. remplissant les conditions requises pour répondre)}}$$

Pour l'ensemble du Manitoba et de l'Alberta, le taux de réponse des enfants membres du panel se chiffre à 98,2 %. Au niveau provincial, les taux de réponse sont :

Manitoba = 2 887 / 2 935 = 98,4 %
Alberta = 3 944 / 4 020 = 98,1 %

Renseignements pertinents pour le calcul des taux de réponse

Nombre de répondants au niveau des ménages :	62 536
Nombre de répondants autres que des enfants au niveau du panel :	59 292
Nombre d'enfants répondants au niveau du panel :	6 831
Nombre de non-répondants au niveau des ménages :	15 645
Nombre de non-répondants autres que des enfants au niveau du panel :	3 244
Nombre d'enfants non répondants au niveau du panel :	124
Nombre de ménages hors du champ de l'enquête :	83 645

Calcul du taux de réponse des ménages :

$$\text{Taux M} = \frac{62\,536}{62\,536 + 15\,645} = \frac{62\,536}{78\,181} = 80{,}0\ \%$$

Calcul du taux de réponse des membres du panel autres que des enfants :

$$\text{Taux MP} = \frac{59\,292}{59\,292 + 3\,244} = \frac{59\,292}{62\,536} = 94{,}8\%$$

Calcul du taux de réponse des enfants membres du panel :

$$\text{Taux MP} = \frac{6\,831}{6\,831 + 124} = \frac{6\,831}{6\,955} = 98{,}2\ \%$$

8.1.3 Taux de réponse transversaux globaux

Taux de réponse des ménages

$$\frac{\textit{nombre de ménages répondants (échantillon principal + échantillons CA)}}{\textit{nombre de ménages dans le champ de l'enquête (échantillon principal + échantillons CA)}}$$

Pour les ménages de l'échantillon principal et des échantillons CA confondus, le taux de réponse transversal se chiffre à ***82,6 %*** au niveau national. Au niveau provincial, le taux varie de 78,8 % en Ontario à 87,3 % au Manitoba, pour les provinces qui ont acquis des unités supplémentaires d'échantillonnage par CA, et de 91,6 % en Colombie-Britannique à 96,5 % à Terre-Neuve pour les autres.

Taux de réponse des membres du panel autres que les enfants choisis par CA

$$\frac{\textit{nombre de répondants autres que les enfants choisis par CA au questionnaire H06 (éch. principal + éch. CA)}}{\textit{nombre de ménages répondants (échantillon principal + échantillons CA)}}$$

Le taux de réponse des membres du panel autres que les enfants choisis par CA est égal à 95,6 % au niveau national et varie de 94,4 % en Ontario à 95,9 % en Alberta pour les provinces qui ont acheté des unités supplémentaires sélectionnées par CA, et de 98,3 % au Québec à 99,3 % au Nouveau-Brunswick pour les autres.

Taux de réponse des enfants choisis par CA membres du panel (2 provinces uniquement)

Soulignons qu'il s'agit du même calcul qu'à la section 8.1.2.

$$\frac{\textit{nombre d'enfants choisis par CA qui ont répondu au questionnaire H06}}{\textit{nombre de ménages répondants comptant un enfant de moins de 12 ans (c.-à-d. remplissant les conditions requises pour répondre)}}$$

Le taux de réponse des enfants choisis par CA membres du panel se chiffre à ***98,2 %*** pour le Manitoba et l'Alberta regroupés. Au niveau provincial, les taux sont :

Manitoba : = 2 887 / 2 935 = 98,4 %
Alberta : = 3 944 / 4 020 = 98,1 %

Renseignements pertinents pour le calcul des taux de réponse

Nombre de répondants au niveau des ménages :	78 751
Nombre de répondants autres que des enfants choisis par CA au niveau du panel :	74 973
Nombre d'enfants choisis par CA répondants au niveau du panel :	6 831
Nombre de non-répondants au niveau des ménages :	16 619
Nombre de non-répondants autres que des enfants choisis par CA au niveau du panel :	3 450
Nombre d'enfants choisis par CA non répondants au niveau du panel :	124
Nombre de ménages hors du champ de l'enquête :	83 732

Calcul du taux de réponse des ménages :

$$\text{Taux M} = \frac{78\,751}{78\,751 + 16\,619} = \frac{78\,751}{95\,370} = 82{,}6\ \%$$

Calcul du taux de réponse des membres du panel autres que les enfants choisis par CA:

$$\text{Taux MP} = \frac{74\,973}{74\,973 + 3{,}450} = \frac{74\,973}{78\,423} = 95{,}6\ \%$$

Calcul du taux de réponse des enfants choisis par CA membres du panel :

$$\text{Taux MP} = \frac{6\,831}{6\,831 + 124} = \frac{6\,831}{6\,955} = 98{,}2\ \%$$

8.2 Erreurs relatives à l'enquête

L'enquête permet de produire des estimations fondées sur les données recueillies auprès d'un échantillon de personnes. 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 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 dépister 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, qu'il n'a pas pu se souvenir du renseignement demandé ou qu'il n'a pas pu fournir de renseignement personnellement ou par personne interposée.

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. 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.

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 :

$$\left(\frac{.003}{.24}\right) x \quad 100 \quad \% \quad ' \quad 1 \ . 25 \quad \%$$

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 fichiers 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 fichiers 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 fichiers 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 progiciels 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 :

SMK6_2 : 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.

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 :

SMK6_4 : 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 :

a) en multipliant la valeur de la variable étudiée par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements étudiés pour obtenir le numérateur ($\hat{X}$),

b) en additionnant les poids finals des enregistrements présentant la caractéristique étudiée pour obtenir le dénominateur ($\hat{Y}$),

c) puis en divisant l'estimation du numérateur par l'estimation du dénominateur.

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 SMK6_4 par le coefficient de pondération WT66, puis on additionne ce résultat pour tous les enregistrements dont la réponse à la question SMK6_2 est «tous les jours» pour obtenir le numérateur ($\hat{X}$). Puis, on additionne les poids finals de tous les enregistrements dont la réponse à la question SMK6_2 est «tous les jours» pour obtenir le dénominateur ($\hat{Y}$). Enfin, on devise ($\hat{X}$) par ($\hat{Y}$) pour obtenir le nombre moyen de cigarettes que fument chaque jour les personnes qui fument tous les jours.

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 progiciels 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 progiciels 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 Tableaux de la variabilité d'échantillonnage approximative (couramment appelées «tableaux de c.v.») pour l'ENSP. On peut employer ces tableaux 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.

Statistique Canada fournit aussi une série de fichiers «bootstrap» et de programmes connexes que les utilisateurs peuvent utiliser à la place des tableaux de c.v. pour calculer des variances individuelles plus précises afin d'évaluer la qualité des totalisations estimatives. De nouveau, on consultera le chapitre 10 pour plus de renseignements.

9.4 Lignes directrices pour la diffusion

Avant de diffuser et/ou de publier des estimations tirées de ces fichiers 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 arrondie et suivre les lignes directrices ci-dessous.

Lignes directrices relatives à la variabilité d'échantillonnage

Type d'estimation	**c.v. (en %)**	**Lignes directrices**
1. Acceptable	0,0 à 16,5	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
2. Marginale	16,6 à 33,3	On peut envisager une diffusion générale non restreinte des estimations, en y joignant une mise en garde aux utilisateurs quant à la variabilité d'échantillonnage élevée liée aux estimations. Les estimations de ce genre doivent être identifiées par la lettre M (ou d'une autre manière similaire).
3. Inacceptable	33,3 ou plus	Statistique Canada recommande de ne pas publier des estimations dont la qualité est inacceptable. Toutefois, si l'utilisateur choisit de le faire, il doit alors adjoindre la lettre I (ou un autre identificateur semblable) et les diffuser avec l'avertissement suivant : «Nous avisons l'utilisateur que ...(précisez les données)... ne répondent pas aux normes de qualité de Statistique Canada pour ce programme statistique. Les conclusions tirées de ces données ne sauraient être fiables et seront fort probablement erronées. Ces données et toute conclusion qu'on pourrait en tirer ne doivent pas être publiées. Si l'utilisateur choisit de les publier, il est alors tenu de publier également le présent avertissement.»

10. Tableaux de la 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 Tableaux de la variabilité d'échantillonnage approximative. Ces «tableaux 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 tableaux 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 tableaux à consulter qui s'appliquerait ensuite à l'ensemble complet des caractéristiques.

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

Données d'entrée des tableaux de la variabilité d'échantillonnage, les provinces et le Canada, membres du ménage, tous les âges

PROVINCE	EFFET DU PLAN	TAILLE DE L'ÉCHAN-TILLON	POPULATION
Terre-Neuve	1,35	3 017	561 586
Île-du-Prince-Édouard	1,53	2 752	135 717
Nouvelle-Écosse	1,41	2 775	917 152
Nouveau-Brunswick	1,34	2 888	745 464
Québec	1,84	7 838	7 221 079
Ontario	1,45	110 845	11 131 489
Manitoba	3,67	32 399	1 085 634
Saskatchewan	1,35	2 785	974 046
Alberta	1,37	40 802	2 728 382
Colombie-Britannique	1,32	4 276	3 780 907
CANADA	5,26	210 377	29 281 456

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

GROUPE D'ÂGE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION
0-11	3,51	37 161	4 686 574
12-24	4,12	39 791	5 134 153
25-44	4,66	66 401	9 709 153
45-64	5,32	44 762	6 335 467
65+	6,01	22 262	3 416 109

Données d'entrée des tableaux de la variabilité d'échantillonnage selon le secteur sanitaire, l'Ontario, le Manitoba et l'Alberta, membres du ménage, tous les âges

SECTEUR SANITAIRE	EFFET DU PLAN	TAILLE DE L'ÉCHAN-TILLON	POPULATION
Ontario			
Ottawa Carleton	1,13	7 143	767 447
Prescott,Russell,Stormont,Dundas,Glen., Renf.	1,19	3 222	287 829
Lanark,Leeds,Gren.,Hast.,P.E.,Fron.,Len.,Add.	1,20	5 247	497 455
Northumberland, Victoria, Haliburton, Peterb.	1,07	4 019	307 511
Durham	1,13	4 914	475 464
Peel	1,09	5 747	893 208
Metro Toronto	1,17	10 978	2 403 746
York	1,13	5 318	626 252
Simcoe	1,12	4 708	348 076
Halton	1,02	4 407	353 830
Niagara	1,02	4 041	418 650
Hamilton-Wentworth	1,09	4 606	484 533

SECTEUR SANITAIRE	EFFET DU PLAN	TAILLE DE L'ÉCHAN-TILLON	POPULATION
Brant, Haldiman, Norfolk	1,12	4 356	226 840
Wellington, Dufferin	1,13	4 504	226 271
Waterloo	1,06	4 393	423 839
Essex	1,08	4 437	354 704
Lambton, Kent	1,02	3 976	242 870
Elgin, Middlesex, Oxford	1,06	4 589	599 688
Bruce, Grey, Perth, Huron	1,03	4 298	301 559
Algoma, Cochrane	1,07	4 124	224 534
Manitoulin, Sudbury	1,00	3 953	205 646
Timiskaming,Muskoka,Parry Sound,Nipissing	1,06	3 852	222 474
Thunder Bay, Kenora, Rainy River	1,05	4 013	239 066
Manitoba			
Eastman Nord et Sud	1,27	6 676	86 532
Burntwood, Norman, Parkland	1,33	6 518	78 881
Central, Interlake	1,33	6 368	153 599
Westman Sud, Brandon, Marquette	1,20	8 501	111 448
Winnipeg	1,12	4 336	655 174
Alberta			
Alberta-Nord	1,23	10 276	418 276
Alberta-Sud	1,14	6 729	335 723
Alberta-Centre	1,14	8 719	427 650
Calgary	1,12	8 236	842 824
Edmonton	1,14	6 842	703 908

Données d'entrée des tableaux de la variabilité d'échantillonnage, les provinces et le Canada, personnes choisies

PROVINCE	ÂGE	EFFET DU PLAN	TAILLE DE L'ÉCHAN-TILLON	POPULATION
Terre-Neuve	2 ans et plus	0,93	963	549 322
Île-du-Prince-Édouard	2 ans et plus	0,93	918	132 322
Nouvelle-Écosse	2 ans et plus	0,94	986	895 914
Nouveau-Brunswick	2 ans et plus	0,92	1 032	728 118
Québec	2 ans et plus	1,11	2 788	7 047 528
Ontario	2 ans et plus	1,15	39 394	10 839 724
Manitoba	Tous	3,71	14 828	1 085 635
Saskatchewan	2 ans et plus	0,93	1 047	948 511
Alberta (excluant les questions de l'EPS)	Tous	1,60	18 305	2 728 383
Colombie-Britannique	2 ans et plus	0,97	1 543	3 686 279
CANADA		4,32	81 804	28 641 735

NOTA: Certaines questions de l'EPS n'ont pas été posées au membres sélectionnés choisis par CA en Alberta. Les variables sont listées à la section 12.1.3. Alors, au moment de calculer des coefficients de variation pour ces variables, il faut utiliser l'âge = 12+, l'effet du plan = 1,16, la taille de l'échantillon = 1 278 et la population = 2 243 982. Un tableau de c.v. est inclus à l'Annexe H pour ces variables.

Données d'entrée des tableaux de la variabilité de l'échantillonnage selon le groupe d'âge, Canada, personnes choisies

GROUPE D'ÂGE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION
2-11 (toutes les provinces)[1]	8,18	7,224	3 925 489
0-11 (Man., Alb.)	2,41	7 114	667 907
2-11 (autres provinces)	1,93	1 288	3 378 942
12-24	5,59	12 120	5 134 153
25-44	4,14	28 900	9 709 158
45-64	3,88	19 019	6 335 468
65+	2,54	13 363	3 416 108

[1] Le groupe d'âge des 2 à 11 ans pour l'ensemble du Canada se retrouve avec l'effet du plan plutôt grand dû au plan d'échantillonnage.

Données d'entrée des tableaux de la variabilité d'échantillonnage selon le secteur sanitaire, l'Ontario, le Manitoba et l'Alberta, personnes choisies

SECTEUR SANITAIRE	ÂGE	EFFET DU PLAN	TAILLE DE L'ÉCHAN-TILLON	POPULATION
Ontario	2 +			
Ottawa Carleton		0,86	2 650	738 276
Prescott,Russell,Stormont,Dundas,Glengarry, Renf.		1,00	1 174	272 703
Lanark,Leeds,Grenville,Hast.,P.E.,Fron.,Len.,Add.		0,90	1 978	486 629
Northumberland,Victoria, Haliburton, Peterb.		1,00	1 513	277 441
Durham		0,72	1 618	482 391
Peel		0,83	1 728	882 317
Metro Toronto		1,05	4 085	2 238 057
York		0,84	1 582	607 098

SECTEUR SANITAIRE	ÂGE	EFFET DU PLAN	TAILLE DE L'ÉCHAN-TILLON	POPULATION
Simcoe		0,82	1 642	339 940
Halton		0,77	1 493	360 691
Niagara		0,85	1 506	406 163
Hamilton-Wentworth		0,76	1 631	484 221
Brant, Haldiman, Norfolk		0,95	1 534	205 422
Wellington, Dufferin		0,63	1 546	250 593
Waterloo		0,90	1 555	383 946
Essex		0,87	1 558	339 105
Lambton, Kent		0,73	1 474	251 089
Elgin, Middlesex, Oxford		0,76	1 688	609 183
Bruce, Grey, Perth, Huron		0,86	1 573	297 485
Algoma, Cochrane		0,68	1 473	237 354
Manitoulin, Sudbury		0,77	1 479	206 157
Timiskaming, Muskoka, Parry Sound, Nipissing		0,70	1 465	233 708
Thunder Bay, Kenora, Rainy River		0,70	1 449	249 755
Manitoba	Tous			
Eastman Nord et Sud		1,29	2 953	86 530
Burntwood, Norman, Parkland		1,28	3 059	78 881
Central, Interlake		1,27	2 803	153 601
Westman Sud, Brandon, Marquette		1,17	4 050	111 449
Winnipeg		1.09	1 963	655 174
Alberta	Tous			
Alberta-Nord		1,36	4 488	418 276

SECTEUR SANITAIRE	ÂGE	EFFET DU PLAN	TAILLE DE L'ÉCHAN-TILLON	POPULATION
Alberta-Sud		1,12	3 070	335 724
Alberta-Centre		1,09	3 927	427 650
Calgary		1,16	3 694	842 824
Edmonton		1,14	3 126	703 909

Tous les coefficients de variation sont *approximatifs* dans les Tableaux de la 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. En outre, il est prévu de fournir des fichiers «bootstrap» et les programmes connexes avec cette diffusion. 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 «inacceptable».

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 tableaux de c.v. pour les estimations de type nominal

Les règles suivantes devraient permettre à l'utilisateur de calculer à partir des Tableaux de la 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 le Tableau 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'un tableau particulier, 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 tableaux, 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 le tableau 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 du tableau) et au numérateur de la proportion ou du pourcentage (côté gauche du tableau). 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 :

$$s_{\hat{d}} = \sqrt{(\hat{X}_1 a_1)^2 + (\hat{X}_2 a_2)^2}$$

où $\hat{x}_1$ représente l'estimation 1, $\hat{x}_2$ l'estimation 2, et a_1 et a_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 $s_{\hat{d}} / \hat{d}$. 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 :

$$s_{\hat{R}} = \hat{R} \sqrt{a_1^2 + a_2^2}$$

où a_1 et a_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 $s_{\hat{R}} / \hat{R} \cdot \sqrt{a_1^2 \% a_2^2}$. 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

Dans ce cas-ci, les règles 3 et 4 sont combinées. 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 tableaux 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-dessus.

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

Supposons qu'un utilisateur estime à 5 794 518 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 au tableau de c.v. pour le CANADA pour les personnes choisies.

2) L'agrégat estimé (5 794 518) 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 1,4 %.

4) Le coefficient de variation approximatif de l'estimation est donc 1,4 %. Par conséquent, l'estimation selon laquelle 5 794 518 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 5 794 518/6 781 835 = 85,4 % 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 au tableau 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 (85,4 %) et la partie numérateur du pourcentage (5 794 518) pour déterminer le coefficient de variation.

3) Le numérateur (5 794 518) 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 en-têtes de colonnes; il faut donc utiliser le nombre qui s'en rapproche le plus, soit 90,0 %.

4) Le nombre qui se trouve à l'intersection de la ligne et de la colonne utilisées, soit 0,5 %, est le coefficient de variation (exprimé en pourcentage) à employer.

5) Le coefficient de variation de l'estimation est donc 0,5 %. Par conséquent, l'estimation selon laquelle 85,4 % 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, 4 902 055/5 794 518 = 85 % 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, 4 261 353/5 692 300 = 75 % 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 SMC6_2, SMC6_4, SMC6_4A, SMC6_5 et SMC6_7. Comment l'utilisateur fait-il pour déterminer le coefficient de variation de la différence entre ces deux estimations?

1) À l'aide du tableau de c.v. pour le CANADA, personnes choisies utilisé de la même façon que dans l'exemple 2, vous établissez à 0,5 % le c.v. de l'estimation 1 (exprimé en pourcentage) et à 1,1 % 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 :

$$s_{\hat{d}} = \sqrt{(\hat{X}_1 a_1)^2 + (\hat{X}_2 a_2)^2}$$

où $\hat{x}_1$ est l'estimation 1, $\hat{x}_2$ est l'estimation 2, et a_1 et a_2 sont les coefficients de variation de $\hat{x}_1$ et $\hat{x}_2$ respectivement.

L'erreur-type de la différence $\hat{d} = (.75 - .85) = .10$ est donc :

$$s_{\hat{d}} = \sqrt{[(.85)(.005)]^2 + [(.75)(.011)]^2}$$
$$= .009$$

3) Le coefficient de variation de $\hat{d}$ est donné par $s_{\hat{d}} / \hat{d} = .009/.10 = 0.095$.

4) Le coefficient de variation approximatif de la différence entre les estimations est donc 9,5 % (exprimé en pourcentage). Par conséquent, cette estimation peut être publiée sans réserve.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 5 794 518 le nombre de personnes qui fument tous les jours et à 987 317 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 au tableau de c.v. pour le CANADA pour les personnes choisies.

3) Le numérateur de cette estimation de rapport est 987 317. Le chiffre qui se rapproche le plus de ce nombre est 1 000 000. Le coefficient de variation de cette estimation (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 3,8 %.

4) Le dénominateur de cette estimation de rapport est 5 794 518. 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 1,4 %.

5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, soit

$$a_{\hat{R}} = \sqrt{8_1^6 \% a_2^2}$$

où a_1 et a_2 sont les coefficients de variation de $\hat{x}_1$ et $\hat{x}_2$ respectivement.

C'est-à-dire,

$$a_{\hat{R}} = \sqrt{(.038)^2 \% (.014)^2}$$

$$= 0.040$$

Le rapport des personnes qui fument occasionnellement à celles qui fument tous les jours est 987 317/5 794 518, soit 0.17:1. 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 tableaux 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 Tableaux de la variabilité d'échantillonnage approximative, en trouvant d'abord dans le tableau approprié le coefficient de variation de l'estimation $\hat{x}$, puis en utilisant la formule suivante pour obtenir l'intervalle de confiance CI correspondant :

$$CI_{\hat{X}} = [\hat{X} - z\hat{X}a_{\hat{X}}, \hat{X} + z\hat{X}a_{\hat{X}}]$$

où $a_{\hat{x}}$ est le coefficient de variation trouvé pour $\hat{x}$, et

z = 1 si l'on désire un intervalle de confiance de 68 %
z = 1.6 si l'on désire un intervalle de confiance de 90 %
z = 2 si l'on désire un intervalle de confiance de 95 %
z = 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 tableaux 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).

$\hat{x} = .854$

z = 2

$a_{\hat{x}} = .005$ est le coefficient de variation de cette estimation selon les tableaux.

$CI_X = \{.854 - (2)(.854)(.005), .854 + (2)(.854)(.005)\}$

$CI_X = \{.846, .863\}$

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

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 $s_{\hat{d}}$.

Si $z = (\hat{X}_1 - \hat{X}_2) / s_{\hat{d}}$ 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 tableaux de c.v. pour effectuer un test Z

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 0.009. Par conséquent,

$$z = \frac{\hat{X}_1 - \hat{X}_2}{s_{\hat{d}}} = \frac{.85 - .75}{.009} = \frac{.10}{.009} = 10.52$$

Puisque z = 10.52 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 Variances ou coefficients de variation réels

Tous les coefficients de variation qui figurent dans les tableaux de la variabilité d'échantillonnage approximative (tableaux de c.v.) sont effectivement approximatifs, donc, non officiels. Toutefois, Statistique Canada peut produire les coefficients de variation réels de variables particulières contre recouvrement des coûts. En outre, il est prévu de fournir des fichiers «bootstrap» et les programmes connexes avec cette diffusion. Les types d'estimation pour lesquels ce service est offert incluent les agrégats, les proportions, les ratios, les différences entre agrégats, 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 bootstrap». 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 tableaux de c.v. publiés 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, les tableaux de c.v. 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 tableaux distincts 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 tels tableaux n'ont pas été produits. 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 tableaux de c.v., mais obtient ainsi un coefficient de variation compris dans la fourchette «marginale» (de 16,6 % à 33,3 %), devrait diffuser les estimations associées en y joignant une mise en garde aux utilisateurs quant à la variabilité d'échantillonnage élevée liée aux estimations. Dans ce cas, il serait bon de recalculer le coefficient de corrélation à l'aide du programme de variance exacte pour vérifier si ces estimations peuvent être diffusées sans mise en garde. Cette situation tient au fait que l'estimation des coefficients de variation grâce aux tableaux de la 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.

On peut se procurer dès maintenant le programme de calcul de la variance ou du coefficient de variation réel et tout utilisateur intéressé par ce service devrait communiquer avec Stéphane Tremblay au (613) 951-4765, Section des méthodes des statistiques sur la santé, Division des méthodes d'enquêtes-ménages, Statistique Canada. Le temps 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 497,52 $; 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. Nous prévoyons également diffuser un deuxième fichier de microdonnées à grande diffusion contenant des facteurs «bootstrap» pour permettre aux utilisateurs de calculer eux-mêmes les coefficients de variation approximatifs pour la plupart des analyses.

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

Les six tableaux suivants indiquent les seuils de diffusion des totaux selon les estimations pour les provinces et le Canada, ainsi que pour les différents groupes d'âges et le secteur sanitaire pour l'Ontario, le Manitoba et l'Alberta, pour les membres du ménage et les personnes choisies. Les estimations basées sur des échantillons de taille inférieure au minimum indiqué dans la colonne «Inacceptables» ne peuvent en aucun cas être diffusées.

Table des seuils de diffusion des totaux selon les estimations, les provinces et le Canada, membres du ménage, tous les âges

PROVINCE	ACCEPTABLES	MARGINALES	INACCEPTABLES
Terre-Neuve	9 000	4 000	2 500
Île-du-Prince-Édouard	2 500	1 000	500
Nouvelle-Écosse	17 000	7 500	4 000
Nouveau-Brunswick	12 500	5 500	3 000
Québec	61 500	27 000	15 500
Ontario	5 500	2 500	1 500
Manitoba	4 500	2 000	1 000
Saskatchewan	17 000	7 500	4 000
Alberta	3 500	1 500	1 000
Colombie-Britannique	42 500	18 500	10 500
CANADA	27 000	11 500	6 500

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

GROUPE D'ÂGE	ACCEPTABLES	MARGINALES	INACCEPTABLES
0-11	16 000	7 000	4 000
12-24	19 500	8 500	5 000
25-44	25 000	11 000	6 000
45-64	27 500	12 000	7 000
65+	33 500	14 500	8 500

Table des seuils de diffusion des totaux selon les estimations selon le secteur sanitaire, l'Ontario, le Manitoba, l'Alberta, membres du ménage, tous les âges

SECTEUR SANITAIRE	ACCEPTABLES	MARGINALES	INACCEPTABLES
Ontario			
Ottawa Carleton	4 500	2 000	1 000
Prescott,Russell,Storm.,Dundas,Glen., Renf.	4 000	1 500	1 000
Lan.,Leeds,Gren.,Hast.,P.E.,Fron.,Len.,Add.	4 000	2 000	1 000
Northumberland,Victoria,Haliburton,Peterb.	3 000	1 500	500
Durham	4 000	1 500	1 000
Peel	6 000	2 500	1 500
Metro Toronto	9 500	4 000	2 500
York	5 000	2 000	1 000
Simcoe	3 000	1 500	500
Halton	3 000	1 500	500
Niagara	4 000	1 500	1 000
Hamilton-Wentworth	4 000	2 000	1 000
Brant, Haldiman, Norfolk	2 000	1 000	500

SECTEUR SANITAIRE	ACCEPTABLES	MARGINALES	INACCEPTABLES
Wellington, Dufferin	2 000	1 000	500
Waterloo	3 500	1 500	1 000
Essex	3 000	1 500	1 000
Lambton, Kent	2 500	1 000	500
Elgin, Middlesex, Oxford	5 000	2 000	1 000
Bruce, Grey, Perth, Huron	2 500	1 000	500
Algoma, Cochrane	2 000	1 000	500
Manitoulin, Sudbury	2 000	1 000	500
Timiskaming,Muskoka,Parry Sound,Nipissing	2 000	1 000	500
Thunder Bay, Kenora, Rainy River	2 500	1 000	500
Manitoba			
Eastman Nord et Sud	500	500	500
Burntwood, Norman, Parkland	500	500	500
Central, Interlake	1 000	500	500
Westman Sud, Brandon, Marquette	500	500	500
Winnipeg	6 000	2 500	1 500
Alberta			
Alberta-Nord	2 000	1 000	500
Alberta-Sud	2 000	1 000	500
Alberta-Centre	2 000	1 000	500
Calgary	4 000	2 000	1 000
Edmonton	4 500	2 000	1 000

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

PROVINCE	ACCEPTABLES	MARGINALES	INACCEPTABLES
Terre-Neuve	19 000	8 500	4 500
Île-du-Prince-Édouard	4 500	2 000	1 000
Nouvelle-Écosse	30 500	13 500	7 500
Nouveau-Brunswick	23 000	10 000	6 000
Québec	101 500	44 500	25 000
Ontario	11 500	5 000	3 000
Manitoba	10 000	4 500	2 500
Saskatchewan	30 000	13 500	7 500
Alberta	8 500	4 000	2 000
Colombie-Britannique	83 000	36 500	21 000
CANADA	55 500	24 000	13 500

Table des seuils de diffusion des totaux selon les estimations des groupes d'âges, Canada, personnes choisies

GROUPE D'ÂGE	ACCEPTABLES	MARGINALES	INACCEPTABLES
2-11 (toutes les provinces)	156 500	70 000	39 500
0-11 (Man., Alb.)	8 000	3 500	2 000
2-11 (autres provinces)	176 500	79 000	45 000
12-24	85 500	37 500	21 500
25-44	51 000	22 000	12 500
45-64	47 000	20 500	11 500
65+	23 500	10 500	6 000

Table des seuils de diffusion des totaux selon les estimations, secteur sanitaire, l'Ontario, le Manitoba, l'Alberta, personnes choisies

SECTEUR SANITAIRE	ACCEPTABLES	MARGINALES	INACCEPTABLES
Ontario			
Ottawa Carleton	8 500	4 000	2 000
Prescott,Russell,Storm.,Dundas,Glen., Renf.	8 500	3 500	2 000
Lan.,Leeds,Gren.,Hast.,P.E.,Fron.,Len.,Add.	8 000	3 500	2 000
Northumberland,Victoria,Haliburton,Peterb.	6 500	3 000	1 500
Durham	8 000	3 500	2 000
Peel	15 500	6 500	4 000
Metro Toronto	21 000	9 000	5 000
York	11 500	5 000	3 000
Simcoe	6 000	2 500	1 500
Halton	6 500	3 000	1 500
Niagara	8 000	3 500	2 000
Hamilton-Wentworth	8 000	3 500	2 000
Brant, Haldiman, Norfolk	4 500	2 000	1 000
Wellington, Dufferin	3 500	1 500	1 000
Waterloo	8 000	3 500	2 000
Essex	7 000	3 000	1 500
Lambton, Kent	4 500	2 000	1 000
Elgin, Middlesex, Oxford	10 000	4 500	2 500
Bruce, Grey, Perth, Huron	6 000	2 500	1 500
Algoma, Cochrane	4 000	1 500	1 000
Manitoulin, Sudbury	4 000	1 500	1 000

SECTEUR SANITAIRE	ACCEPTABLES	MARGINALES	INACCEPTABLES
Timiskaming,Muskoka,Parry Sound,Nipis.	4 000	2 000	1 000
Thunder Bay, Kenora, Rainy River	4 500	2 000	1 000
Manitoba			
Eastman Nord et Sud	1 500	500	500
Burntwood, Norman, Parkland	1 000	500	500
Central, Interlake	2 500	1 000	500
Westman Sud, Brandon, Marquette	1 000	500	500
Winnipeg	13 000	6 000	3 500
Alberta			
Alberta-Nord	4 500	2 000	1 000
Alberta-Sud	4 500	2 000	1 000
Alberta-Centre	4 500	2 000	1 000
Calgary	9 500	4 000	2 500
Edmonton	9 500	4 000	2 500

11. Pondération

La composante des ménages de l'Enquête nationale sur la santé de la population de 1994-1995 comportait deux plans de base : un pour les neufs provinces à part le Québec et un autre pour le Québec. Pour les neuf provinces à part le Québec, l'ENSP se fondait sur le plan de l'Enquête sur la population active (EPA), avec de nombreuses modifications, pour produire un échantillon qui lui est propre. Pour cette raison, le calcul des facteurs de pondération est relié à la méthode de pondération appliquée à l'EPA. En revanche, pour le Québec, on a suivi un plan d'échantillonnage à deux phases; l'échantillon de première phase a été prélevé en 1992-1993 par l'Enquête sociale et de santé (ESS) et celui de deuxième phase, par l'ENSP. Par conséquent, au Québec, le calcul des facteurs de pondération est relié à la méthode de pondération employée dans l'ESS. Pour plus de détails, voir le chapitre 5 - Plan d'échantillonnage. En 1996-1997, on a prélevé un petit échantillon supplémentaire indépendant par composition aléatoire (CA) en Ontario, au Manitoba et en Alberta, afin de pouvoir produire des estimations fiables au niveau du secteur sanitaire.

Pour le premier cycle, deux ensembles de facteurs de pondération étaient nécessaires, l'un pour les estimations transversales fondées sur les variables de la composante générale à laquelle ont répondu tous les membres des ménages en 1994-1995 et l'autre, pour les estimations transversales fondées sur les variables de la composante santé à laquelle a répondu uniquement le membre sélectionné comme répondant longitudinal. Pour le deuxième cycle, on s'est servi de trois ensembles de facteurs de pondération pour calculer les données transversales. Les deux premiers sont semblables à ceux de 1994-1995, mais sont calculés différemment pour tenir compte du plan d'échantillonnage de 1996-1997.

Le troisième facteur de pondération, qui ne s'applique qu'au fichier de données sur la santé, doit être utilisé pour l'analyse de populations et de variables particulières. Pour plus de renseignements, consulter le chapitre 12.

La méthode de pondération de 1996-1997 se fonde en grande partie sur celle de 1994-1995. On prend comme point de départ les facteurs de pondération finals de 1994-1995 appliqués à chaque répondant longitudinal qui continue à participer à l'enquête. Certaines corrections, qui ne sont plus nécessaires en 1996-1997, ont été supprimées, pour créer un facteur de pondération «épuré» pour chaque répondant longitudinal. À partir de là, on apporte les nouvelles corrections pour produire les facteurs de pondération finals de 1996-1997, pour le membre du ménage choisi comme répondant longitudinal (membre du panel) et pour chaque membre courant du ménage. Les sections qui suivent donnent une brève description des méthodes de pondération de 1994-1995 qui restent applicables pour la pondération de 1996-1997. On trouvera une description complète des méthodes appliquées en 1994-1995 dans le Guide du fichier de microdonnées à grande diffusion de l'Enquête nationale sur la santé de la population de 1994-1995. Les nouvelles corrections qu'il faut apporter en 1996-1997 en ce qui concerne l'échantillon principal de l'ENSP sont décrites à la section 11.1, tandis que les

méthodes appliquées en 1996-1997 aux données sur les provinces qui ont acquis des unités d'échantillonnage additionnels par CA (Ontario, Manitoba et Alberta) sont exposées à la section 11.2. La méthode de pondération pour les provinces sauf le Québec est décrite à la section 11.3 et celle pour le Québec, à la section 11.4.

11.1 Pondération de l'échantillon transversal de l'ENSP de 1996-1997 — Échantillon principal des ménages

On décrit à la présente section les méthodes de pondération appliquées aux membres du panel et à tous les membres de leur ménage inclus dans l'échantillon principal de l'ENSP. On trouvera à la section 11.2 une description complète des méthodes de pondération supplémentaire qu'il a fallu utiliser pour tenir compte des achats d'unité d'échantillonnage supplémentaire de l'Ontario, du Manitoba et de l'Alberta.

11.1.1 Facteurs de pondération épurés

Comme on l'a décrit aux sections 11.3 et 11.4, le point de départ de la méthode de pondération de 1996-1997 est le facteur de pondération «épuré», calculé d'après le plan d'échantillonnage original de 1994-1995. Une fois ces facteurs de pondération calculés, on effectue les corrections décrites ci-après.

11.1.2 Corrections des facteurs de pondération appliqués aux membres des ménages

Correction 1A : Correction pour la non-réponse des ménages

Par définition, la catégorie des ménages non répondants englobe les ménages dont la non-réponse est due à un refus, à des circonstances spéciales, à un obstacle linguistique, au fait que le logement est inoccupé, à une absence temporaire ou à un problème informatique.

Dans certains cas, le membre du ménage sélectionné pour faire partie du panel en 1996-1997 était décédé ou placé en établissement, avait déménagé au Yukon ou dans les Territoires du Nord-Ouest, ou avait quitté le pays.

Aux fins de la pondération transversale, ces ménages sont considérés comme des ménages répondants à ce stade, mais sont éliminés des calculs subséquents. Ces unités ne figurent pas dans le fichier de microdonnées transversales, mais sont enregistrées dans le fichier longitudinal.

Pour tenir compte des ménages entiers qui n'ont pas répondu à l'enquête de 1996-1997, on effectue la correction suivante :

$$\frac{\textit{somme des poids pour les ménages répondants et non répondants dans la catégorie de pondération}}{\textit{somme des poids pour les ménages répondants dans la catégorie de pondération}}$$

Les catégories de pondération représentent des groupes d'unités (ou de ménages) qui affichent la même propension à répondre à l'enquête. On se fonde sur les caractéristiques des répondants et des non-répondants au deuxième cycle de l'enquête obtenue d'après les données du premier cycle pour déterminer à quelle catégorie de pondération ils appartiennent. On définit les catégories à l'aide d'un algorithme de mise en grappe qui permet de disposer les unités d'échantillonnage en une structure arborescente en divisant successivement les ensembles de données en «branches» d'après les caractéristiques des unités. Chaque division a pour objet de répartir les unités existantes en au moins deux groupes le plus dissemblables possible en ce qui a trait aux taux observés de non-réponse (et à l'intérieur desquels on s'attend à observer des taux de non-réponse fort similaires). On peut se servir d'une autre caractéristique pour définir chaque division. Par exemple, on peut d'abord répartir les unités en un groupe de logements occupés par le propriétaire et un groupe de logements loués. Puis, on peut subdiviser le premier groupe en cinq sous-groupes, d'après le niveau de revenu du ménage, tandis qu'on peut subdiviser le deuxième groupe d'après l'âge du répondant. Chaque groupe nouvellement formé peut ensuite être divisé, d'après d'autres caractéristiques, et ainsi de suite. Les résultats des subdivisions finales correspondent aux catégories de pondération.

On se sert du logiciel *Knowledge Seeker IV for Windows*, développé par ANGOSS Software International Limited, pour produire la structure arborescente. Nous nous sommes servis de la version améliorée de l'algorithme CHAID (Chi-Square Automatic Interaction Detector) offerte par *Knowledge Seeker* pour repérer, à chaque nœud, la caractéristique qui permet le mieux de diviser l'échantillon en deux groupes dissemblables en regard d'une caractéristique donnée, ici l'indicateur de réponse/non-réponse. Quand on se sert de données nominales comptant plus de deux niveaux, *Knowledge Seeker* peut regrouper un ou plusieurs niveaux, de sorte que le nombre de «branches» soit inférieur au nombre de catégories. Pour les caractéristiques continues, comme l'âge, *Knowledge Seeker* commence par subdiviser les données en dix intervalles, que l'on peut ou non regrouper, de façon à n'obtenir parfois que deux «branches». À chaque étape, on effectue des tests statistiques pour s'assurer de ne produire que des subdivisions statistiquement significatives. On apporte des corrections de Bonferroni au niveau de signification des divers tests, pour s'assurer d'atteindre le niveau de signification de chaque division. La subdivision

se termine quand on ne peut plus effectuer aucune division statistiquement significative ou quand la division produit des catégories trop petites (chaque catégorie doit contenir au moins 30 unités). Pour plus de renseignements sur l'algorithme CHAID, consulter Kass (1980)[3].

On se fonde sur les caractéristiques personnelles du membre du ménage sélectionné pour faire partie du panel, ainsi que sur celles du logement ou du ménage, pour définir les catégories de pondération pour la non-réponse des ménages. On estime que les caractéristiques personnelles des répondants longitudinaux jouent un rôle important dans la prédiction de la non-réponse d'un ménage pour plusieurs raisons. Souvent, la non-réponse est la même au niveau de la personne qu'au niveau du ménage. Le fait que l'impossibilité de dépister certains répondants longitudinaux en 1996-1997 a mené à la non-réponse du ménage en est un exemple manifeste. En outre, durant la collecte des données, on a insisté sur le fait d'obtenir une réponse du membre du panel, puisque les renseignements sur ce dernier sont essentiels à l'exécution de l'étude longitudinale. Si le répondant longitudinal n'était pas disponible ou refusait de répondre, les intervieweurs avaient l'instruction de ne pas demander aux autres membres du ménage de répondre à la composante générale. Enfin, dans nombre de ménages, les membres présentaient des caractéristiques communes, comme la race. Le cas échéant, les caractéristiques du répondant sont dans une certaine mesure, celles du ménage.

Des ensembles distincts de cellules de correction de la pondération sont créés pour chaque province. En plus des caractéristiques du ménage et des caractéristiques personnelles du membre choisi pour faire partie du panel, on s'est servi de certaines caractéristiques liées au plan d'enquête pour limiter l'effet du plan de sondage sur les résultats de l'inférence statistique et pour intégrer les variables de conception de l'enquête à l'analyse. Les caractéristiques varient selon la province, mais englobent les variables suivantes :

Géographiques	Province, région métropolitaine de recensement, indicateur de région urbaine/rurale
Ménage	Type de logement, situation de propriétaire/locataire, type de famille, caractère adéquat du revenu du ménage, source principale de revenu, indicateur de non-réponse aux questions sur le revenu en 1994-1995, présence d'enfants dans le ménage.
Personnel	Sexe, âge, indicateur d'âge supérieur à 16 ans, état civil, race, pays de naissance, âge au moment de l'immigration, indicateur de limitation de l'activité, activité principale/activité sur le marché du travail.

Correction 2A : Migrations interprovinciales

Il est parfois nécessaire de faire une correction pour tenir compte des membres du panel qui déménagent d'une province très peuplée à une province peu peuplée, car les facteurs de pondérations appliqués à ces membres sont souvent anormalement grands comparativement à ceux appliqués à d'autres unités similaires dans la nouvelle province. On n'apporte une correction que si on observe au moins un facteur de pondération extrême dans le schéma de migration interprovinciale observé. Le cas échéant, on diminue les facteurs de pondération appliqués à tous les membres du panel qui répondent à ce schéma de migration, de sorte que la somme de ces facteurs soit égale à la projection démographique du nombre de personnes ayant déménagé au cours des deux dernières années dans le cadre du schéma de migration observé. Si la correction se traduit par une augmentation du facteur de pondération, on ne la fait pas, puisqu'elle accroîtrait encore davantage le facteur de pondération extrême.

Correction 3A : Méthode du partage des facteurs de pondération

Seuls les membres du panel (répondants longitudinaux) sont suivis d'un cycle à l'autre de l'enquête. Donc, si la composition du ménage du répondant longitudinal varie entre deux cycles, aucun facteur de pondération n'aura été associé aux cohabitants observés au deuxième cycle, puisque ces derniers n'avaient pas été sélectionnés dans le panel au premier cycle. La méthode du partage des facteurs de pondération est un mécanisme qui permet d'attribuer des facteurs de pondération à ces personnes de façon à produire des estimations non biaisées. Typiquement, on attribue à chaque membre du ménage le facteur de pondération du membre du panel divisé par le nombre de membres du ménage qui faisaient partie du champ d'observation de l'enquête au premier cycle (par exemple, en excluant les personnes nées au Canada depuis 1994 ou qui y sont entrées depuis cette date). Pour plus de renseignements, consulter Ernst (1989)[2] ou Lavallée (1995)[4].

Correction 4A : Correction pour la non-réponse des membres des ménages

Pour le deuxième cycle, le taux de non-réponse attribuable à un membre du ménage est inférieur à 2 %. Par conséquent, on combine tous les groupes âge-sexe et on fait la correction qui suit, par province :

$$\frac{\textit{somme des facteurs de pondération des répondants et des non-répondants dans une province}}{\textit{somme des facteurs de pondération des répondants dans une province}}$$

Correction 5A : Intégration de l'échantillon principal et des échantillons supplémentaires

Comme on l'a mentionné précédemment, au deuxième cycle, l'Ontario, le Manitoba et l'Alberta ont effectué des achats importants d'unités supplémentaires sélectionnées par composition aléatoire (CA). Par conséquent, pour ces trois provinces, il existe des échantillons provenant de deux bases de sondage, la base de sondage ordinaire (principale) de l'ENSP et la base de CA. Il est nécessaire de combiner les facteurs de pondération des unités principales et des unités sélectionnées par CA afin de ne pas produire des estimations en double de la population. On applique à cette fin une méthode à base de sondage double, qui est une adaptation de la méthode de Skinner et Rao (1996)[5]. Cette correction est décrite en détail à la section 11.4.

Correction 6A : Correction par stratification a posteriori

Cette correction vise à assurer que la somme des facteurs de pondération finals corresponde aux totaux de population de 1996-1997 au niveau province-âge-sexe. Les catégories âge-sexe regroupent les personnes de 0 à 11 ans, de 12 à 24 ans, de 25 à 44 ans, de 45 à 64 ans et de 65 ans et plus, hommes et femmes. La correction est donnée par

$$\frac{\textit{projection démographique dans une catégorie province-âge-sexe}}{\textit{somme des poids des membres des ménages répondants dans une catégorie province-âge-sexe}}$$

De nouveau, il convient de souligner que les ménages composés entièrement d'immigrants arrivés en 1994 ou après ne sont pas couverts par l'échantillon principal de l'ENSP de 1996-1997, mais sont inclus dans les projections démographiques. Par conséquent, on traite implicitement ces immigrants comme si leurs caractéristiques étaient semblables à celles du reste de la population.

Correction 7A : Facteur de bruit

Pour des raisons de confidentialité, on ajoute un facteur de «bruit» aux facteurs de pondération des personnes qui appartiennent à un même ménage. On choisit ce facteur, qui obéit à une loi de distribution uniforme, de façon que la somme des facteurs de pondération au niveau du ménage ne varie pas.

Le facteur de pondération final appliqué à chaque membre du ménage dans le fichier général est calculé en multipliant le facteur de pondération «épuré» par toutes les corrections susmentionnées.

11.1.3 Corrections de la pondération pour les membres du panel

Pour calculer le facteur de pondération final appliqué aux membres du panel dans le fichier de données sur la santé, on multiplie le facteur de pondération «épuré» par les corrections qui suivent. (À noter que, pour faciliter les comparaisons, on a adopté ici la même numérotation des corrections qu'à la section précédente.)

Correction 1B : Correction pour la non-réponse des membres du ménage et du membre du panel

Cette correction tient compte de la non-réponse du ménage entier, ainsi que de celle de certains membres du ménage répondant qui ne répondent pas au questionnaire destiné au membre du panel. De la même façon que pour la correction 1A, on construit des catégories de pondération en se servant des renseignements dont on dispose sur les membres du panel sélectionnés en 1994-1995. De nouveau, on crée des ensembles distincts de cellules de correction pour les diverses provinces, mais les données d'entrée utilisées pour définir les catégories sont légèrement différentes. Ces catégories sont basées sur les données provenant de tous les répondants longitudinaux, plutôt qu'uniquement de ceux qui ont répondu au moins à la composante générale les deux années, comme c'est le cas pour le calcul de la correction pour la non-réponse des membres du ménage. La correction est donnée par :

$$\frac{\textit{somme des poids pour les ménages répondants et non répondants dans la catégorie de pondération}}{\textit{somme des poids pour les ménages répondants dans la catégorie de pondération}}$$

Correction 2B : Migration interprovinciales

Cette correction est similaire à la correction 2A. Dans les cas où on observe un facteur de pondération extrême pour le schéma de migration interprovinciale observé, on réduit les facteurs de pondération appliqués à tous les membres du panel visés par le schéma de migration, de façon que leur somme soit égale à la projection démographique du nombre de personnes ayant déménagé au cours des deux dernières années conformément au schéma de migration visé. Si la correction se traduit par une augmentation du facteur de pondération, on ne l'applique pas, puisqu'elle augmentera encore davantage la valeur du facteur de pondération extrême.

Correction 5B : Intégration des échantillons supplémentaires sélectionnés par CA

Comme la correction 5A, on effectue cette correction pour combiner les facteurs de pondération des unités principales et des unités sélectionnées par CA pour éviter de produire des estimations en double de la population. Voir la section 11.2 pour plus de précisions.

Correction 6B : Stratification a posteriori

Cette correction assure que la somme des facteurs de pondération finals qui figurent dans le fichier sur la santé soit égale aux totaux de population de 1996-1997 au niveau province-âge-sexe. Les catégories âge-sexe comptent les personnes de 2 à 11 ans, de 12 à 24 ans, de 25 à 44 ans, de 45 à 64 ans et de 65 ans et plus — hommes et femmes. Cette correction est donnée par

$$\frac{\textit{projection démographique dans une catégorie province-âge-sexe}}{\textit{somme des facteurs de pondération des répondants individuels dans une catégorie province-âge-sexe}}$$

Puisqu'on n'a sélectionné aucun nouveau répondant longitudinal en 1996-1997, le premier groupe d'âge ne couvre que les personnes de 2 à 11 ans. De nouveau, il convient de souligner que les ménages composés entièrement d'immigrants arrivés en 1994 ou après ne sont pas couverts par l'échantillon principal de l'ENSP de 1996-1997, mais sont inclus dans les projections démographiques. Par conséquent, on traite implicitement ces immigrants comme s'ils avaient les mêmes caractéristiques que le reste de la population.

11.2 Pondération transversale de l'échantillon de l'ENSP de 1996-1997 — Provinces qui ont acheté un échantillon supplémentaire sélectionné par CA

Introduction

Pour tirer pleinement parti des renseignements transversaux que fournissent les données, il est nécessaire de regrouper les données tirées de l'échantillon principal de l'ENSP et celles provenant des échantillons supplémentaires sélectionnés par CA en un grand ensemble de données et de calculer des facteurs de pondération qui reflètent ce regroupement des sources de données. Dans certaines provinces où des unités sont sélectionnées par CA, on pose des questions supplémentaires aux répondants choisis de cette façon, ou bien on pose certaines questions à d'autres groupes d'âge que dans le cas de l'enquête principale. Les écarts entre les populations couvertes par ces questions obligent à calculer des facteurs de pondération supplémentaires. En tout, on calcule trois facteurs de pondération transversaux, un pour l'analyse du fichier général et deux pour l'analyse du fichier sur la santé. On se sert du *facteur de pondération final type pour les membres du ménage* (WT56) pour analyser les données provenant du questionnaire général.

Le *facteur de pondération final type pour les membres «adultes» du panel* et le *facteur de pondération final type pour les enfants membres du panel* ont été regroupés pour donner le facteur de pondération WT66. On se sert de ce facteur de pondération pour analyser la plupart des données du fichier sur la santé obtenues pour les personnes de

12 ans et plus et pour les enfants de 0 à 11 ans sélectionnés au Manitoba et en Alberta. Le deuxième facteur de pondération qui figure dans le fichier sur la santé, WT66_N, peut être appliqué à deux populations distinctes et devrait donc être utilisé avec prudence. On se sert notamment de ce facteur de pondération (*facteur de pondération final de l'EPS pour les membres «adultes» du panel*) pour analyser les données du questionnaire sur la santé englobant toute question qui faisait partie de l'Enquête sur la promotion de la santé. Il s'applique aux personnes de 12 ans et plus. On se sert aussi de ce facteur de pondération pour analyser les questions sur les services de santé destinés aux enfants (SVB6_1 à SVB6_5). Ces questions n'ont été posées qu'aux enfants faisant partie des échantillons sélectionnés par CA en Alberta et au Manitoba.

Consulter la section 12.1 intitulée «Utilisation des facteurs de pondération» pour plus de précision sur les conditions d'utilisation des divers facteurs de pondération.

Vue d'ensemble

La production de facteurs de pondération dans le cas des provinces qui ont acheté des unités supplémentaires sélectionnées par CA comprend quatre étapes générales. Premièrement, on produit les facteurs de pondération principaux jusqu'à l'étape de la stratification a posteriori, sans inclure cette dernière (pour d'autres précisions, voir la section précédente). Puis, on définit les facteurs de pondération des unités sélectionnées par CA jusqu'à une étape similaire. Les calculs pour les unités sélectionnées par CA sont passés en revue dans la présente section. Ensuite, on intègre les deux ensembles de facteurs de pondération pour produire un ensemble unique en appliquant une méthode mise au point par Skinner et Rao (1996)[5]. Enfin, on effectue la stratification a posteriori de ces facteurs de pondération conformément aux totaux de population, au niveau secteur sanitaire-âge-sexe.

Méthode d'intégration de Skinner et Rao

On applique cette méthode pour combiner deux échantillons indépendants et pondérés individuellement, comme l'échantillon principal et les échantillons sélectionnés par CA de l'ENPS, afin d'obtenir un échantillon unique auquel s'applique un seul ensemble de facteurs de pondération composés. On détermine un facteur a (0#a#1) de façon à indiquer l'importance relative de chaque échantillon dans le fichier de données combinées. On multiplie le facteur de pondération de l'échantillon principal par a, tandis qu'on multiplie celui des échantillons sélectionnés par CA par 1-a. Cette méthode s'applique si l'échantillon principal et les échantillons supplémentaires sélectionnés par CA couvrent la même population, ce qui n'est pas toujours le cas pour l'ENSP. À titre d'exemple, mentionnons les enfants d'un an et moins qui répondaient aux critères pour faire l'objet d'un questionnaire détaillé dans la composante de sélection par CA, mais pas

dans la composante principale. Dans ce cas, il est nécessaire d'appliquer la méthode de façon plus poussée.

On détermine la valeur a non pas au niveau provincial, mais à celui appelé niveau agrégé du secteur sanitaire. On divise les provinces en plusieurs grandes régions géographiques, souvent en séparant les grandes villes et les régions relevant des bureaux régionaux. En Alberta, on distingue trois secteurs sanitaires agrégés, à savoir Calgary, Edmonton et le reste de la province. Au Manitoba, il en existe deux, à savoir Winnipeg et le reste de la province. En Ontario, les secteurs sont au nombre de six — Toronto, Ottawa-Carleton et quatre autres régions géographiques. On se sert de régions agrégées de façon que la taille des échantillons régionaux soit suffisante pour produire des valeurs d'entrée stable pour le calcul d'a, tout en permettant de tenir compte des écarts qui pourraient survenir entre des régions distinctes (grandes villes par opposition au reste de la province). Pour plus de renseignements sur la méthode proprement dite, consulter Skinner et Rao (1996)[5].

Calcul des facteurs de pondération des unités sélectionnées par CA

Les méthodes utilisées pour calculer les facteurs de pondération des unités sélectionnées par CA sont similaires à celles appliquées à l'échantillon principal. On produit un facteur de pondération initial en se fondant sur la probabilité de sélectionner le numéro de téléphone en question. Puis, on apporte plusieurs corrections pour tenir compte des numéros sélectionnés qui n'ont pas été envoyés aux bureaux régionaux ou pour tenir compte d'autres événements durant la collecte des données, comme l'existence de ménages possédant plusieurs lignes téléphoniques. Dans de nombreux cas, on définit les corrections sous forme de ratios pondérés, en se servant des facteurs de pondération calculés à l'étape précédente pour déterminer l'importance de la correction. On obtient alors le facteur de pondération résultant en multipliant le facteur précédent par le facteur de correction.

11.2.1 Facteurs de pondération de base des unités sélectionnées par CA

La composition aléatoire a été mise en oeuvre selon la méthode de l'élimination des banques inactives qui produit un échantillon aléatoire simple stratifié (sans remise) des lignes téléphoniques résidentielles. On a subdivisé la totalité de la province en plusieurs strates de CA (consulter le Chapitre 5 sur le plan d'échantillonnage pour plus de précisions), qui ne correspondaient pas aux strates utilisées pour le plan d'échantillonnage principal dans la province. Chaque mois, on a sélectionné un échantillon de numéros de téléphone à partir de la base de sondage pour la CA. Le facteur de pondération mensuel initial correspond à l'inverse de la probabilité de sélectionner le numéro de téléphone particulier d'après la liste des banques actives de numéro durant ce mois. Les

facteurs de pondération mensuels sont convertis en facteurs de pondération de base globaux en multipliant, au niveau de la strate, les facteurs de pondération pour le mois *i* par

$$\frac{\textit{taille de l'échantillon de la strate pour le mois } i}{\textit{taille totale de l'échantillon de la strate}}$$

Si les taux d'échantillonnage ou la base de sondage (la liste des banques actives) ne varient pas pendant l'enquête, alors tous les facteurs de pondération sont égaux dans la strate. Si un changement a lieu, on tient compte de l'augmentation ou de la diminution du taux d'échantillonnage ou du nombre de banques actives en appliquant un poids plus ou moins élevé aux enregistrements effectués pour le mois en question.

11.2.2 Corrections supplémentaires des facteurs de pondération de base

Correction 1 : Sous-échantillonnage du Manitoba

Pour chaque strate, on définit un «taux de succès» théorique. Ce taux correspond au nombre prévu de numéros de téléphone résidentiels dans la strate (tel que déterminé d'après un fichier de factures de téléphone) comparativement au nombre total possible de numéros de téléphone dans la strate après l'application de la méthode de l'élimination des banques inactives. Au Manitoba, le taux de succès a été très faible pour plusieurs strates. Par conséquent, on a tiré un grand échantillon dont la plupart des numéros étaient inactifs, ce qui représentait un échantillon fort peu efficace. Pour ces régions, on a décidé d'écarter environ la moitié des numéros échantillonnés qui ne figuraient pas parmi les numéros résidentiels du fichier de facturation. (L'élimination de toutes les unités sélectionnées mènerait à l'exclusion des nouveaux numéros et des numéros non publiés, donc pourrait introduire un biais.) Par conséquent, on applique le facteur multiplicatif :

$$\frac{\textit{échantillon total de numéros de téléphone non résidentiels dans la strate}}{\textit{échantillon de numéros de téléphone non résidentiels dans la strate envoyés au bureau régional}}$$

aux numéros non résidentiels de la strate qui ont été envoyés aux bureaux régionaux aux fins de l'interview. Ce facteur est toujours pratiquement égal à 2.

Correction 2 : Numéros de téléphone sélectionnés deux fois

Puisque l'enquête principale et les enquêtes par CA couvrent la même population, mais que les échantillons ont été sélectionnés indépendamment, il est possible qu'un ménage ait été sélectionné pour les deux composantes. Pour éviter autant que possible cette situation, on a comparé les numéros de téléphone sélectionnés par CA à ceux figurant dans la base de données de l'échantillon principal, pour voir s'il existait des numéros en double. En cas d'appariement, le numéro sélectionné par CA n'a pas été envoyé au bureau régional. Or, il faut tenir compte de ces ménages pour calculer les facteurs de pondération appliqués aux numéros qui ont été envoyés aux bureaux régionaux. On multiplie ces facteurs de pondération par un facteur égal à

$$\frac{\textit{somme des facteurs de pondération de tous les ménages}}{\textit{somme des facteurs de pondération des enregistrements envoyés aux bureaux régionaux}}$$

On effectue cette correction au niveau de la strate.

Correction 3 : Correction pour la non-réponse des ménages

Pour tenir compte des ménages entiers qui n'ont pas répondu, on fait la correction suivante :

$$\frac{\textit{somme des facteurs de pondération pour les ménages répondants et non-répondants}}{\textit{somme des facteurs de pondération pour les ménages répondants}}$$

De nouveau, on effectue cette correction au niveau de la strate. Si la valeur du facteur de correction est élevée (supérieure à 2,5), la strate en cause est regroupée avec les strates voisines jusqu'à ce que la correction tombe sous le seuil de 2,5.

Correction 4 : Correction pour les ménages sans téléphone

Manifestement, les personnes qui n'ont pas le téléphone dans leur logement n'ont aucune chance d'être sélectionnées et interviewées dans le cadre de l'enquête par CA. Si on se contente d'ignorer ce facteur, on introduit un biais en regard des personnes ayant des caractéristiques semblables à celles qui ne possèdent pas de téléphone. Ciok (1993)[1] a étudié la population qui ne possède pas de téléphone au Canada. Il a réparti la population en cinq catégories de ménages et a examiné dans chacune la population ne possédant pas de téléphone.

Les cinq catégories sont :
i) ménages d'une seule personne — résidents de 65 ans et plus;
ii) ménages de deux personnes ou plus — tous les résidents de 65 ans et plus;
iii) ménages d'une seule personne — résidents de moins de 65 ans;
iv) ménages monoparentaux;
v) autres ménages.

D'après les données de cette étude, ainsi que les pourcentages provinciaux de ménages sans téléphone, on a calculé des facteurs de correction pour chaque province au niveau des cinq groupes définis par Ciok. Puis, on a multiplié les facteurs de pondération par le facteur de correction pour les ménages sans téléphone.

Correction 5 : Lignes de téléphone multiples

Les logements possédant plus d'une ligne téléphonique résidentielle ont une probabilité plus forte de faire partie de l'échantillon sélectionné par CA. On recueille durant l'interview les renseignements sur le nombre de numéros de téléphone résidentiels que possède le ménage. Le facteur de pondération résultant des corrections précédentes est alors multiplié par l'inverse du nombre de lignes résidentielles. Donc, on diviserait par deux le facteur de pondération d'un ménage qui possède deux lignes résidentielles.

Les facteurs de pondération résultant de cette correction sont appelés **«facteurs de pondération démographiques de l'échantillon sélectionné par CA»**.

11.2.3 Corrections supplémentaires des facteurs de pondération pour les membres du ménage

Lors du calcul des facteurs de pondération démographiques appliqués aux unités sélectionnés par CA, on a effectué toutes les corrections d'après les caractéristiques au niveau du ménage ou du logement. À partir de ce stade, les corrections sont apportées au niveau des personnes. Les corrections décrites dans la présente section ont trait aux facteurs de pondération appliqués aux membres du ménage, autrement dit les facteurs de pondération utilisés quand on se sert des données provenant de la composante générale à laquelle ont répondu tous les membres du ménage.

Correction 1H : Logements collectifs

Les logements collectifs, dans lesquels vivent ensemble au moins 10 personnes non apparentées, font partie du champ d'observation de l'enquête principale, mais non de la composante de la CA. Puisque les deux volets sont regroupés ultérieurement, il est important que les facteurs de pondération calculés pour la composante de la CA reflètent également la population des logements collectifs. Comme dans le cas de la population ne possédant pas de téléphone, si on n'applique qu'un seul facteur de correction à l'ensemble de la population, on produira un biais. Afin d'obtenir un meilleur facteur de correction, on a extrait des renseignements sur les résidents des logements collectifs des fichiers du Recensement de 1991. On a subdivisé la population en 10 catégories âge-sexe (hommes et femmes, de 0 à 11 ans, de 12 à 24 ans, de 25 à 44 ans, de 45 à 64 ans et de 65 ans et plus) et en trois catégories d'état civil (marié(e), célibataire, divorcé(e)/séparé(e)/veuf(ve)). Dans chacune de ces catégories, on a calculé la proportion de personnes vivant dans un logement collectif. À noter que certains logements correspondant, aux termes du recensement, à un logement collectif, comme les hôpitaux, les prisons et les bases militaires, ne font pas partie du champ d'observation de l'ENSP. Seules les catégories communes au recensement et à l'ENSP sont prises en compte dans le calcul du nombre de résidents des logements collectifs. Dans chaque province, les poids démographiques appliqués pour la CA à chaque catégorie âge-sexe-état civil ont été multipliés par le facteur

(1 + *proportion de personnes dans la catégorie âge-sexe-état civil dans les logements collectifs*).

Correction 2H : Correction pour la non-réponse des membres du ménage

Cette correction permet de tenir compte des membres des ménages répondants qui ne répondent pas au questionnaire à l'intention des membres du ménage. La correction est égale à

$$\frac{\textit{somme des facteurs de pondération des répondants et non-répondants individuels dans une catégorie âge-sexe}}{\textit{somme des facteurs de pondération des répondants individuels dans une catégorie âge-sexe}}$$

Les catégories âge-sexe comprennent les personnes de 0 à 11 ans, de 12 à 24 ans, de 25 à 44 ans, de 45 à 64 ans et de 65 ans et plus — hommes et femmes. Les corrections sont faites au niveau de chaque secteur sanitaire. On appelle le facteur de pondération obtenu après cette correction **«facteur de pondération du membre du ménage pour la CA»**.

Intégration des membres des ménages de l'échantillon principal et des échantillons sélectionnés par CA

Facteur de pondération type

On se sert de ce facteur de pondération quand on analyse les réponses aux questions de la composante générale. Puisque l'échantillon principal ainsi que les échantillons sélectionnés par CA couvrent tous les intervalles d'âge, on applique une version simplifiée de la méthode de Skinner et Rao pour combiner les facteurs de pondération des membres des ménages de l'échantillon principal et les facteurs de pondération des membres des ménages des échantillons obtenus par CA. Le facteur de pondération résultant est appelé **«facteur de pondération intégré type du membre du ménage»**. L'étape finale du calcul du facteur de pondération type est la stratification a posteriori.

Correction 3HS : Stratification a posteriori

On dispose de projections démographiques mensuelles pour plusieurs catégories âge-sexe à divers niveaux géographiques. Elles sont fondées sur les données du recensement le plus récent ainsi que sur les estimations de la natalité, de la mortalité, de l'immigration et de l'émigration. On se sert de ces données pour déterminer les totaux au niveau secteur sanitaire-âge-sexe dans les trois provinces qui ont acheté des unités supplémentaires, puis, on corrige les facteurs de pondération en les multipliant par le facteur suivant :

$$\frac{\textit{projection démographique dans une catégorie secteur sanitaire-âge-sexe}}{\textit{somme des facteurs de pondération des membres des ménages répondants dans une catégorie secteur sanitaire-âge-sexe}}$$

Les catégories âge-sexe sont les mêmes que celles utilisées pour calculer la correction pour la non-réponse des membres du ménage, autrement dit les personnes de 0 à 11 ans, de 12 à 24 ans, de 25 à 44 ans, de 45 à 64 ans et de 65 ans et plus — hommes et femmes. Le facteur de pondération résultant est appelé **«facteur de pondération final type du membre du ménage» (WT56)**.

11.2.4 Corrections supplémentaires des facteurs de pondération pour les membres du panel

Pour le calcul des facteurs de pondération démographiques des échantillons sélectionnés par CA, toutes les corrections tenaient compte des caractéristiques du ménage ou du logement. À partir d'ici, les corrections sont faites au niveau de la personne. Les corrections décrites dans la présente section ont trait aux facteurs de pondération appliqués aux membres du panel, autrement dit les facteurs de pondération utilisés quand on se sert des données provenant de la composante santé à laquelle ne répondent que les membres «adultes» du panel, âgés de 12 ans et plus, ou les enfants membres du panel dans le cas des échantillons supplémentaires sélectionnés par CA en Alberta et au Manitoba (rappelons que, dans ces provinces, on sélectionne à la fois un membre de 12 ans et plus et un enfant de moins de 12 ans).

Correction 1S : Probabilité de sélection dans le ménage

Tous les membres de 12 ans et plus d'un ménage répondant sélectionné par CA avaient la même chance d'être choisis comme membre «adulte» du panel. En outre, en Alberta et au Manitoba, tous les enfants de moins de 12 ans d'un ménage avaient la même chance d'être sélectionnés (on ne pouvait sélectionner les enfants en Ontario). On doit donc augmenter les facteurs de pondération appliqués aux membres du panel pour tenir compte de cette étape supplémentaire de sélection. Ce facteur de correction multiplicatif est égal à l'inverse de la probabilité de sélection. Pour les personnes de 12 ans et plus sélectionnées, cela signifie que la correction est égale au nombre de personnes de 12 ans et plus dans le ménage, tandis que pour les enfants, elle est égale au nombre d'enfants dans le ménage.

Correction 2S : Logements collectifs

Cette correction est appliquée de la même façon que la correction 1H, mais, dans ce cas-ci, on utilise les facteurs de pondération pour les membres du panel.

Correction 3S : Correction pour la non-réponse des membres du panel

Cette correction permet de tenir compte des membres des ménages répondants sélectionnés pour faire partie du panel qui ne répondent pas au questionnaire destiné aux membres du panel. La correction est égale à

$$\frac{\textit{somme des facteurs de pondération des répondants et non-répondants individuels dans une catégorie âge-sexe}}{\textit{somme des facteurs de pondération des répondants individuels dans une catégorie âge-sexe}}$$

Les catégories âge-sexe comprennent les personnes de 0 à 11 ans, de 12 à 24 ans, de 25 à 44 ans, de 45 à 64 ans et de 65 ans et plus — hommes et femmes. Les corrections sont calculées pour chaque secteur sanitaire. Après cette correction, le facteur de pondération résultant est appelé **«facteur de pondération pour les membres du panel sélectionné par CA»**.

Intégration des membres du panel de l'échantillon principal et des échantillons sélectionnés par CA

Deux facteurs de pondération sont nécessaires pour les membres du panel. Pour les membres de 12 ans et plus, on applique un facteur de pondération type à la plupart des données d'analyse et on se sert d'un facteur de pondération de l'Enquête sur la promotion de la santé (EPS) pour certaines questions qui n'ont pas été posées aux membres de l'échantillon sélectionné par CA en Alberta. Pour les provinces qui n'ont pas acheté d'unités d'échantillonnage supplémentaire, le facteur de pondération type et le facteur de l'EPS sont identiques. En ce qui concerne les enfants, on se sert aussi d'un facteur de pondération type dans la plupart des cas, ainsi que d'un facteur de pondération visant les services de santé pour certaines questions sur les services de santé auxquels ont accès les enfants qui n'ont été posées qu'aux membres des échantillons sélectionnés par CA en Alberta et au Manitoba. Comme on l'a mentionné précédemment, les facteurs de pondération finals types pour les enfants et les «adultes» membres du panel ont été regroupés en une seule zone du fichier de microdonnées (WT66). Le facteur de pondération final de l'EPS pour les membres «adultes» du panel et le facteur de pondération final visant les services de santé pour les enfants membres du panel ont été regroupés pour former la zone WT66_N dans le fichier de microdonnées. Dans tous les cas, on se sert d'une version de la méthode d'intégration de Skinner et Rao pour combiner le facteur de pondération des membres du panel provenant de l'échantillon principal et le facteur de pondération des membres du panel provenant des échantillons sélectionnés par CA. Enfin, on procède à une stratification a posteriori pour déterminer le facteur de pondération final.

Facteur de pondération type du membre «adulte» du panel

On se sert de ce facteur de pondération pour analyser la plupart des réponses de la composante santé. Puisque tant l'échantillon principal que l'échantillon sélectionné par CA couvre l'intervalle d'âge de 12 ans et plus, on se sert de la version simplifiée de la méthode de Skinner et Rao pour produire un ensemble de facteurs de pondération appelés **«facteur de pondération intégré type pour les membres «adultes» (de 12 ans et plus) du panel»**.

Correction 3SSA : Stratification a posteriori

On dispose de projections démographiques mensuelles pour plusieurs catégories âge-sexe à divers niveaux géographiques. Elles sont fondées sur les données du recensement le plus récent, ainsi que sur les estimations de la natalité, de la mortalité, de l'immigration et de l'émigration. On se sert de ces données pour calculer les totaux de population au niveau secteur sanitaire-âge-sexe pour les trois provinces qui ont acheté des unités d'échantillonnage supplémentaire. Puis, on corrige les facteurs de pondération en les multipliant par le facteur

$$\frac{\textit{projection démographique dans une catégorie secteur sanitaire-âge-sexe}}{\textit{somme des facteurs de pondération pour les membres des ménages répondants dans une catégorie secteur sanitaire-âge-sexe}}$$

Les catégories âge-sexe sont les mêmes que celles utilisées pour apporter la correction pour la non-réponse des membres, autrement dit les personnes de 12 à 24 ans, de 25 à 44 ans, de 45 à 64 ans et de 65 ans et plus — hommes et femmes. Le facteur de pondération résultant est appelé **«facteur de pondération final type pour les membres «adultes» (de 12 ans et plus) du panel» (WT66)**.

Facteur de pondération de l'EPS

On se sert de ce facteur de pondération pour analyser les réponses à certaines questions de l'Enquête sur la promotion de la santé. L'Alberta a décidé de ne pas poser ces questions aux membres de l'échantillon supplémentaire sélectionné par CA. Par conséquent, pour cette province, les échantillons principal et sélectionné par CA ne sont pas intégrés. Le facteur de pondération pour les membres du panel provenant de l'échantillon principal est stratifié a posteriori de la même façon que celui calculé pour les autres provinces qui n'ont pas sélectionné d'échantillon supplémentaire par CA (voir la section 11.3.3). Il n'est pas stratifié a posteriori au niveau du secteur sanitaire,

à cause de la petitesse de l'échantillon. Au Manitoba et en Ontario, le facteur de pondération est le même que le facteur de pondération final type pour les membres du ménage. Ce facteur de pondération est appelé **«facteur de pondération final de l'EPS pour les membres «adultes» (de 12 ans et plus) du panel» (WT66_N)**.

Facteur de pondération type pour les enfants

On se sert de ce facteur de pondération quand on analyse les réponses à la plupart des questions supplémentaires posées à l'enfant qui est membre du panel. Puisque l'échantillon principal ne compte aucun enfant de 0 ou de 1 an, il s'agit d'un cas spécial de l'application de la méthode d'intégration de Skinner et Rao. Pour ces deux provinces, on regroupe les répondants de 2 à 11 ans des deux enquêtes. Pour ceux de 1 an ou moins (0), on ne possède que les enregistrements obtenus par CA. La méthode prévoit aussi un dénombrement de population qui permet de s'assurer de la représentativité des deux sous-groupes (groupe des 0 et 1 an et des 2 à 11 ans). Le facteur de pondération résultant est appelé **«facteur de pondération intégré type pour les enfants du panel»**. Il n'est pas nécessaire de le calculer pour l'Ontario, puisqu'on n'a sélectionné aucun enfant dans l'échantillon obtenu par CA. Donc, on peut calculer les facteurs de pondération pour l'Ontario de la même façon que pour toute province où aucun échantillon n'est sélectionné par CA et, pareillement, les stratifier a posteriori au niveau de la province plutôt qu'à celui du secteur sanitaire.

Correction 3SSC : Stratification a posteriori

On dispose de projections démographiques mensuelles pour plusieurs catégories âge-sexe à divers niveaux géographiques. Elles sont fondées sur les données du recensement le plus récent, ainsi que sur les estimations de la natalité, de la mortalité, de l'immigration et de l'émigration. On se sert de ces données pour calculer les totaux de population au niveau secteur sanitaire-âge-sexe pour les trois provinces qui achètent des unités d'échantillonnage supplémentaires. Puis, on corrige les facteurs de pondération en les multipliant par le facteur

$$\frac{\textit{projection démographique dans une catégorie secteur sanitaire-âge-sexe}}{\textit{somme des facteurs de pondération pour les membres des ménages répondants dans une catégorie secteur sanitaire-âge-sexe}}$$

Les catégories âge-sexe sont les mêmes que celles utilisées pour calculer la correction pour la non-réponse des membres, autrement dit les garçons et les filles de 0 à 11 ans. Le facteur de pondération résultant est appelé **«facteur de pondération final type pour les enfants du panel» (WT66)**.

Facteur de pondération pour les services de santé destinés aux enfants

On se sert de ce facteur de pondération pour analyser les réponses aux questions sur les services de santé offerts aux enfants. Ces questions n'ont été posées qu'aux enfants sélectionnés dans les échantillons choisis par CA de l'Alberta et du Manitoba. Il n'existe pas de données provenant de l'échantillon principal auxquelles les intégrer. En dehors de ces deux provinces, le facteur de pondération est nul.

Correction 3SHC : Stratification a posteriori

On dispose de projections démographiques mensuelles pour plusieurs catégories âge-sexe à divers niveaux géographiques. Elles sont fondées sur les données du recensement le plus récent, ainsi que sur les estimations de la natalité, de la mortalité, de l'immigration et de l'émigration. On se sert de ces données pour calculer les totaux de population au niveau secteur sanitaire-âge-sexe pour les trois provinces qui ont acheté des unités d'échantillonnage supplémentaires. Puis, on corrige les facteurs de pondération en les multipliant par le facteur

$$\frac{\textit{projection démographique dans une catégorie secteur sanitaire-âge-sexe}}{\textit{somme des facteurs de pondération pour les membres des ménages répondants dans une catégorie secteur sanitaire-âge-sexe}}$$

Les catégories âge-sexe sont les mêmes que celles utilisées pour calculer les corrections pour la non-réponse des membres, autrement dit les garçons et les filles de 0 à 11 ans. Le facteur de pondération résultant est appelé **«facteur de pondération final des services de santé pour les enfants du panel» (WT66_N)**.

11.3 Méthodes de pondération fondées sur celles de 1994-1995 pour les provinces à part le Québec

On présente d'abord une brève description de la méthode de pondération fondamentale de l'EPA. Puis, on décrit les autres corrections apportées aux facteurs de pondération multiplicatifs pour calculer les facteurs de pondération «épurés».

11.3.1 Facteurs de pondération de base de l'EPA

L'EPA s'appuie sur un plan d'échantillonnage à plusieurs degrés stratifié (essentiellement à deux degrés mais, dans certains cas, à trois). D'une façon ou de l'autre, on sélectionne un échantillon de grappes dans chaque strate selon l'une des méthodes d'échantillonnage avec probabilité proportionnelle à la taille (PPT). Puis, on calcule un «facteur de pondération de la grappe» de l'EPA, qui représente la probabilité inverse de sélection d'une grappe, conformément à la méthode d'échantillonnage choisie. Au degré suivant (dernier degré), on sélectionne des logements dans les grappes échantillonnées selon une méthode d'échantillonnage systématique. On calcule un «facteur de pondération du logement», qui représente la probabilité inverse de sélection d'un logement, compte tenu du fait que la grappe qui le contient a été sélectionnée. On attribue ensuite un «facteur de pondération de base de l'EPA» en multipliant le facteur de pondération de la grappe par celui du logement.

11.3.2 Corrections supplémentaires des facteurs de pondération de base

Correction 1 : Correction pour le groupe de renouvellement

L'échantillon de l'EPA se compose de six «groupes de renouvellement». Pour l'ENSP, on prélève un échantillon de l'EPA en choisissant un nombre entier (de 1 à 6) de groupes de renouvellement. Il arrive cependant qu'on doive sélectionner un nombre fractionnaire pour obtenir la taille d'échantillon nécessaire (voir correction 3).

Par conséquent, la première correction effectuée pour compenser l'effet du nombre entier est la suivante :

$$\frac{\textit{nombre de renouvellements dans une strate de l'EPA utilisé pour l'EPA (habituellement 6)}}{\textit{nombre entier de renouvellements dans une strate de l'EPA nécessaire pour l'ENSP}}$$

Correction 2 : Correction pour l'expansion de la grappe

Il se peut que la taille de la grappe augmente entre le moment du recensement et celui 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 en cas de croissance modérée du parc immobilier. Si les grappes subissent une forte croissance, on procède à un sous-échantillonnage pour que la tâche des intervieweurs ne soit pas trop lourde. L'ENSP prévoit un sous-échantillonnage analogue des grappes qui affichent une croissance modérée.

Donc, le second facteur de correction multiplicatif correspond à l'application du rapport inverse de sous-échantillonnage aux grappes sous-échantillonnées dans le cadre de l'EPA ou 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. Le «secteur de stabilisation» comprend les grappes de la base de sondage des ménages à revenu élevé et des ménages vivant en appartement, ainsi que des groupes de 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» par sous-échantillonnage la fraction du groupe de renouvellement de l'EPA dont on n'a pas besoin pour l'ENSP (voir Correction pour le groupe de renouvellement). La troisième correction s'effectue comme suit :

$$\frac{\textit{nombre de logements sélectionnés dans une grappe pour l'EPA}}{\textit{nombre de logements d'une grappe effectivement utilisés pour l'ENSP}}$$

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 n'interroge que les occupants d'un seul logement (choisi au hasard), on gonfle le facteur de pondération associé au logement. La quatrième correction correspond 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.

Correction 5 : Correction pour la non-réponse du ménage

Malgré tous les efforts des intervieweurs, il est inévitable que certains ménages ne répondent pas. La non-réponse peut être due à un refus, à des circonstances particulières, à un obstacle linguistique, à l'inoccupation du logement, à une absence temporaire ou à un problème d'ordinateur. On compense les non-

réponses en corrigeant proportionnellement les facteurs de pondération des ménages répondants.

C'est aussi à cette étape qu'on tient compte de l'élimination des unités supplémentaires transversales achetées en 1994-1995, mais qui n'ont pas été étudiés en 1996-1997. On retranche les ménages sélectionnés en 1994-1995 dans le cadre de l'achat d'unités supplémentaires du numérateur ainsi que du dénominateur du calcul qui suit. La cinquième correction s'effectue comme suit :

$$\frac{\textit{somme des facteurs de pondération des ménages échantillonnés pour le groupe strate-saison de l'ENSP}}{\textit{somme des facteurs de pondération des ménages répondants pour le groupe strate-saison de l'ENSP}}$$

Notons qu'on procède à cette correction chaque «saison», c'est-à-dire l'été et l'hiver, 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. La correction est calculée séparément pour chaque saison car le taux de non-réponse varie significativement d'une saison à l'autre. 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 pour chaque groupe strate-saison.

Correction 6 : Correction pour la méthode d'exclusion

Tel qu'indiqué au chapitre 5 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 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. Pour l'Île-du-Prince-Édouard, la correction est légèrement différente, notamment parce que la méthode d'exclusion est appliquée aux quatre périodes de collecte des données plutôt qu'aux deux derniers trimestres. Remarquons aussi qu'on n'applique pas 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 pour ces trois strates.

11.3.3 Corrections supplémentaires pour les membres du panel

Les données du questionnaire du membre du panel viennent d'une seule personne du ménage échantillonné. Le facteur de pondération «épuré» associé à chaque sujet est calculé de la façon suivante. On multiplie d'abord les facteurs de pondération de base de l'EPA par les facteurs de correction 1 à 6, ainsi que 7B, 8B et 9B décrits plus bas. Si la personne sélectionnée pour faire partie du panel est un enfant de moins de 12 ans qui vit dans un logement de l'échantillon «enfants» (voir le chapitre 5 - Plan d'échantillonnage pour la définition d'un logement de l'échantillon «enfants»), en 1994-1995, on a demandé à tous les enfants du ménage, jusqu'à concurrence de quatre, de répondre au questionnaire de l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ). Autrement, on a demandé au membre du panel de répondre à un ensemble supplémentaire de questions de l'ENSP.

On doit apporter plusieurs corrections pour tenir compte de ce plan d'échantillonnage et de la non-réponse à ce questionnaire.

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

En 1994-1995, au cours des deux derniers trimestres de collecte des données de l'ENSP, on a sélectionné des répondants pour les questionnaires destinés aux membres du panel de l'ENSP et de l'ELNEJ. Dans les ménages «enfants» échantillonnés qui avaient des enfants, on a sélectionné jusqu'à quatre enfants (de moins de 12 ans) pour répondre aux questions de l'ELNEJ. Un des enfants était désigné comme étant celui qui serait le membre du panel pour ce ménage à l'occasion des cycles suivants. Les donnés sur les enfants ne figuraient pas dans le fichier de microdonnées de 1994-1995 de l'ENSP.

Pour plus de renseignements sur l'intégration de l'ELNEJ, consulter le chapitre 5 - Plan d'échantillonnage. En 1996-1997, tous les membres du panel, y compris les enfants qui avaient moins de 12 ans en 1994-1995, ont été interrogés dans le cadre de l'ENSP, mais non dans celui de l'ELNEJ.

En 1994-1995, les ménages comptant des enfants de 12 ans et plus ont été tirés de l'échantillon «adultes» uniquement. Pour tenir compte du fait que les ménages de l'échantillon «enfants» n'ont pas contribué aux estimations concernant les membres du panel en 1994-1995, mais le feront en 1996-1997, on corrige les facteurs de pondération applicables aux membres des ménages avec enfants tirés des deux derniers trimestres de l'échantillon «adultes» en les multipliant par l'inverse de la proportion 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, on apporte une correction distincte à des groupes de strates de l'EPA (correspondant habituellement aux strates de l'ENSP), afin d'assurer la concordance 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 chaque ménage échantillonné devient membre du panel. Cette sélection nécessite une correction, qui correspond à la probabilité inverse de sélection. Au départ, on avait l'intention 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 d'exécution de l'IAO a empêché de sélectionner des enfants de 12 ans au cours des deux premiers trimestres. En compensation, au lieu d'appliquer la même probabilité de sélection à chaque membre du ménage durant les deux derniers trimestres, on a augmenté celle 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.

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 des deux premiers trimestres, où les probabilités d'être retenu étaient nulles. Cette correction est appliquée aux groupes de strates de l'EPA qui correspondent habituellement aux strates de l'ENSP, sauf 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» chaque trimestre, mais ils n'ont en fait été sélectionnés dans cet échantillon qu'au cours des deux derniers. Puisque les strates de l'ENSP comptaient, pour la plupart, 40 % de l'échantillon «adultes» au cours des deux derniers trimestres, on a gonflé le facteur de pondération appliqué aux enfants de 12 ans sélectionnés durant ce laps de temps en le multipliant par l'inverse de ce 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, à cause de l'erreur susmentionnée. Donc, on a multiplié le facteur de pondération appliqué aux enfants de 12 ans dans les ménages avec jeunes mais sans enfants 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, pour la strate des ménages vivant en appartement, pour celle des ménages à revenu élevé et pour celle des ménages habitant dans une région éloignée.

Les «facteurs de pondération» susmentionnés sont les facteurs de pondération de base de l'EPA multipliés par tous les facteurs de correction calculés jusqu'à présent (c'est-à-dire les corrections 1 à 6, ainsi que 7B, 8B et 9B).

Ces facteurs de pondération, qui sont les facteurs de pondération «épurés» de 1996-1997 pour les provinces sauf le Québec, sont le point de départ de la pondération de 1996-1997.

11.4 Méthodes de pondération pour le Québec fondées sur celles de 1994-1995

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 le chapitre 5 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 sections 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 «épurés» de 1996-1997 applicables aux participants à l'ENSP.

11.4.1 Facteurs de pondération de l'ESS

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

Facteur de pondération appliqué aux grappes de l'ESS

L'ESS est fondée sur 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 mesure de taille retenue était le nombre de ménages dans la grappe, selon le Recensement de 1986. Le «**facteur de pondération de la grappe de l'ESS**» correspond à la probabilité inverse de sélectionner la grappe en question.

Facteur de pondération appliqué aux logements 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 du logement de l'ESS**» correspond donc à la probabilité inverse de sélectionner le logement dans la grappe, multipliée par le facteur de pondération de la grappe de l'ESS.

11.4.2 Facteur de pondération de base appliqué aux logements de l'ENSP

L'échantillon de l'ENSP a été sélectionné 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 retenue de l'ESS.

Probabilité de retenir 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 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 d'être retenue).

Probabilité de retenir 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 le champ de l'ESS (entreprises, institutions, habitations démolies ou logements vacants) avaient une probabilité de rétention égale à un, au cas où ils feraient partie du champ d'observation de l'ENSP. On a d'abord tiré, parmi les logements couverts par l'ESS, 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 variait de 1/3 pour les ménages d'une seule personne à 1 pour les ménages avec enfants de moins de 12 ans.

Le «**facteur de pondération de base du logement**» correspond au facteur de pondération du logement de l'ESS, multiplié par l'inverse du produit de la probabilité de rétention de la grappe de l'ESS par la probabilité de rétention du logement de l'ESS. La probabilité de rétention du logement de l'ESS inclut la probabilité que le logement soit sélectionné pour l'ENSP au départ et la probabilité qu'il soit retenu en raison de la composition du ménage.

11.4.3 Corrections supplémentaires des facteurs de pondération de base

Correction 1 : 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 2 : 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 :

$$\frac{\textit{numération de l'ENSP}}{\textit{numération de l'ESS}}$$

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 :

$$\frac{\textit{nombre de logements supplémentaires listés}}{\textit{nombre de logements supplémentaires sélectionnés}}$$

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 3 : Correction pour la non-réponse du ménage

Pour tenir compte du nombre total de ménages non répondants, on procède à la correction suivante :

$$\frac{\textit{somme des facteurs de pondération des ménages répondants et non répondants}}{\textit{somme des facteurs de pondération des ménages non répondants}}$$

Ici, le facteur de pondération correspond au facteur de pondération préliminaire.

Une autre correction s'impose au niveau du secteur de pondération pour la non-réponse. En ce qui concerne les logements couverts par l'ESS, le secteur de pondération pour la non-réponse correspond à l'intersection entre une strate de l'ENSP et un type de ménage de l'ESS, par trimestre. 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 champ 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 (collectif 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.

11.4.4 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 le chapitre 5 sur le plan d'échantillonnage pour la définition de l'échantillon «enfants»), tous les enfants du ménage, jusqu'à concurrence de quatre, ont été priés de répondre au questionnaire de l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ). Dans les autres cas, on a posé au membre du panel (de 12 ans et plus) la série de questions supplémentaires de l'ENSP. Plusieurs corrections ont dû être apportées pour tenir compte de ce plan d'échantillonnage et de la non-réponse à ce questionnaire.

Correction 4 : Correction pour l'intégration de l'ELNEJ

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. Les données sur ces enfants ne figuraient pas dans le fichier de microdonnées de l'ENSP de 1994-1995. Il faut 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 les logements de l'échantillon «adultes» où des enfants ont été découverts à l'occasion de l'ENSP.

La correction correspond à l'inverse du taux de sous-échantillonnage pour l'échantillon «adultes» et varie selon qu'on avait ou non observé des enfants dans le logement au moment de l'ESS et selon la catégorie de densité urbaine de l'ESS dans laquelle se classe le logement.

Les logements où des enfants ont été observés au moment de l'ESS et les logements sans enfants sont l'objet d'une correction distincte, car le taux de sous-échantillonnage est différent pour ces deux catégories. Pour Montréal et pour les capitales régionales, la correction 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.

Correction 5 : Probabilité inverse de sélection des membres du panel

En principe, 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 devait avoir, au départ, la même chance d'être sélectionné pour faire partie du panel. Une erreur dans le logiciel a malheureusement empêché la sélection des enfants de 12 ans au cours des deux premiers trimestres. Pour compenser cette erreur, on a appliqué aux enfants de 12 ans une probabilité de sélection égale au double de la précédente durant les troisième et quatrième trimestres. Le facteur de correction appliqué à un membre donné du ménage correspond donc à la probabilité inverse qu'il soit sélectionné comme membre du panel.

Correction 6 : Correction pour les enfants de 12 ans

Afin que les enfants de 12 ans soient représentés correctement, on a dû augmenter le facteur de pondération pour tenir compte des ménages où les enfants de cet âge n'avaient pu être sélectionnés à cause du problème logiciel. La correction correspond à la probabilité inverse qu'un enfant de 12 ans satisfasse les critères de sélection dans un logement 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 seul trimestre. Or, les enfants de 12 ans ne pouvaient être retenus durant les trimestres 1 et 2. Si on veut tenir compte de cette situation, la correction doit s'effectuer au niveau de la strate de l'ENSP au lieu de la grappe. Par souci de 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.

On calcule le facteur de pondération «épuré» de 1996-1997 en multipliant le facteur de pondération démographique par tous les facteurs de correction mentionnés dans la présente section.

Bibliographie

[1] Ciok, R. (1993). The Non-telephone Population—A Preliminary Study. Internal Document. Statistics Canada.

[2] Ernst, L. (1989). Weighting Issues for Household and Family Estimates. In *Panel Surveys.* (Eds. D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh.) New York: John Wiley and Sons, 135-159.

[3] Kass, G.V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 29, 119-127.

[4] Lavallée, P. (1995). Cross-sectional Weighting of Longitudinal Surveys of Individuals and Households Using the Weight Share Method. *Survey Methodology*, 21, 25-32.

[5] Skinner, C.J. and Rao, J.N.K. (1996). Estimation in Dual Frame Surveys with Complex Designs. *Journal of the American Statistical Association*, 91, 433, 349-356.

12. Utilisation du fichier

La présente section débute par un examen des variables de pondération et fournit des explications sur la façon de les utiliser quand on effectue des totalisations d'après les fichiers de microdonnées à grande diffusion. Puis, on y explique la convention appliquée, pour tous les cycles de l'ENSP, pour nommer les variables. Enfin, on y décrit diverses méthodes d'accès aux données que peuvent adopter les analystes.

12.1 Utilisation des facteurs de pondération

12.1.1 Facteur de pondération transversal - Fichier général - WT56

Un seul facteur de pondération, WT56, figure dans le fichier général. Ce facteur de pondération s'applique à tous les groupes d'âge et à toutes les provinces. Il est fondé sur les échantillons principal et sélectionné par CA intégrés. ON DEVRAIT SE SERVIR DE CE FACTEUR DE PONDÉRATION POUR ANALYSER TOUTES LES QUESTIONS AYANT TRAIT AU FICHIER GÉNÉRAL.

(Pour des renseignements détaillés sur le calcul de ce facteur de pondération, consulter les sections 11.1 et 11.2 du chapitre sur la pondération.)

12.1.2 Facteur de pondération transversal - Fichier sur la santé - WT66

Le facteur WT66 s'applique à tous les groupes et à toutes les provinces, mais la façon dont il couvre le groupe d'âges les plus jeunes (de 0 à 11 ans) varie selon la province. Au Manitoba et en Alberta, dans les ménages choisis par CA comptant des enfants, on a sélectionné un enfant (de 0 à 11 ans) en plus d'une autre personne qu'un enfant (de 12 ans et plus) pour répondre à la composante santé. La couverture du facteur WT66 englobe les enfants de 0 à 11 ans. Dans les autres provinces, seuls les enfants du panel sont suivis en 1996-1997, de sorte que le groupe d'âge auquel le facteur s'applique est celui des 2 à 11 ans. Ce facteur de pondération se fonde sur les échantillons principal et sélectionné par CA intégrés. WT66 EST LA PRINCIPALE VARIABLE DE PONDÉRATION DANS LE FICHIER SUR LA SANTÉ ET DEVRAIT ÊTRE UTILISÉE POUR L'ANALYSE DE LA PLUPART DES VARIABLES. Les exceptions, auxquelles il faut appliquer le facteur WT66_N, sont décrites au point 12.1.3.

(Pour une explication plus détaillée du calcul de ce facteur de pondération, consulter les sections 11.1 et 11.2 du chapitre sur la pondération.)

12.1.3 Facteur de pondération transversal - Fichier sur la santé - WT66_N

On peut se servir du facteur WT66 pour analyser les réponses à la plupart des questions du fichier sur la santé. Cependant, il faut se servir du facteur WT66_N pour analyser les réponses aux questions de l'Enquête sur la promotion de la santé (EPS) ainsi qu'aux questions sur les services de santé destinés aux enfants énumérées ci-après.

Variables à analyser en appliquant le facteur WT66_N :

Questions de l'Enquête sur la promotion de la santé
GHS6_11, GHS6_12, GHS6_13, GHS6_14, GHS6_15, GHS6_16A, GHS6_16B, GHS6_16C, GHS6_16D, GHS6_16E, GHS6_16F, GHS6_16G, GHS6_16H, GHS6_17, GHS6_18A, GHS6_18B, GHS6_18C, GHS6_18D, GHS6_18E, GHS6_18F, GHS6_18G, GHS6_18H, GHS6_18I
GHC6_21
HWS6_4, HWS6_5
HVS6_1, HVS6_2, HVS6_3, HVS6_4, HVS6_5, HVS6_6, HVS6_7, HVS6_8, HVS6_9
SMS6_8, SMS6_9, SMS6_13A, SMS6_13B, SMS6_13C, SMS6_13D, SMS6_13E, SMS6_13F, SMS6_13G, SMS6_13H, SMS6_14, SMS6_15, SMS6_16A, SMS6_16B, SMS6_16C, SMS6_16D, SMS6_17, SMS6_18A, SMS6_18B, SMS6_18C, SMS6_18D
ALS6_1, ALS6_2, ALS6_3, ALS6_4, ALS6_5, ALS6_6, ALS6_7
SHS6_4, SHS6_5, SHS6_6, SHS6_7, SHS6_7A
Questions sur les services de santé destinés aux enfants
SVB6_1, SVB6_2, SVB6_3, SVB6_4, SVB6_5

Les questions de l'EPS n'ont pas été posées aux unités échantillonnées par CA en Alberta et s'appliquent uniquement aux répondants «adultes» (12 ans et plus). Les questions sur les services de santé offerts aux enfants ont été posées uniquement aux enfants sélectionnés par CA au Manitoba et en Alberta.

(Pour une explication plus détaillée des composantes de pondération individuelles utilisées pour calculer cette variable de pondération, consulter la section 11.2.4 du chapitre sur la pondération, plus précisément les rubriques «Facteur de pondération de l'EPS», «Facteur de pondération type pour les enfants» et «Facteur de pondération pour les services de santé destinés aux enfants».)

12.2 Convention appliquée pour nommer les variables

En 1996-1997, on a adopté pour nommer les variables de l'ENSP une convention qui permet aux utilisateurs des données de faire facilement référence à des données similaires provenant d'autres périodes de collecte et aux diverses composantes d'enquête du programme de l'ENSP. Les exigences qui suivent devaient être satisfaites : limiter les noms des variables à huit caractères au plus pour qu'il soit facile de les utiliser avec les logiciels d'analyse; préciser la période de référence de l'enquête (1994-1995, 1996-1997, 1998-1999...) dans le nom et permettre de repérer facilement les variables conceptuellement identiques d'une période de référence à l'autre. Par exemple, des données conceptuellement identiques sur l'usage du tabac ont été collectées en 1994-1995 et en 1996-1997. Les noms des variables sur l'usage du tabac ne devraient différer qu'en ce qui concerne la position réservée dans le nom à l'année de référence particulière durant laquelle les données ont été collectées. Cette convention sera appliquée pendant toute la durée de l'enquête longitudinale et sera adoptée pour toutes les composantes de l'ENSP, à savoir la composante des ménages, la composante des établissements de soins, la composante des résidents du Nord et les suppléments.

12.2.1 Structure élémentaire des noms des variables

Chacun des huit caractères du nom d'une variable fournit des renseignements sur le type de données que contient la variable.

Positions 1 et 2 :	Nom de la variable / section du questionnaire
Position 3 :	Type d'enquête
Position 4 :	Année/cycle durant lequel apparaît la variable
Position 5 :	Type de variable
Positions 6 à 8 :	Numéro/nom de la variable tiré du questionnaire

Par exemple :
la signification des noms de variable DHC4_AGE et DHC6_AGE est la suivante :

DH : dans la section du questionnaire sur la démographie et la composition du ménage;
C : questions qui font partie de la composante principale de l'enquête auprès des ménages;
4. figure dans le cycle de 1994-1995;
6 : figure dans le cycle de 1996-1997;
_ : peut être trouvé sur le questionnaire;
AGE : nom de la variable.

12.2.2 Positions 1 et 2 : Nom de la variable / section du questionnaire

On se sert des valeurs suivantes pour le nom de la section de l'enquête :

AD	Dépendance à l'égard de l'alcool	IJ	Blessures
AL	Alcool	IN	Revenu
AM	Administration (de l'enquête)	IS	Assurance
AP	Opinions au sujet des parents	LF	Population active
BP	Tension artérielle	MH	Santé mentale
CC	Problèmes de santé chroniques	PA	Activités physiques
CE	Abandon du contact	PC	Examen général
CI	Renseignement sur le contact (établissements, 1996-1997)	PR	Province
CO	Adaptation (échantillon supplémentaire de l'Alberta, 1994-1995 et 1996-1997)	PY	Ressources psychologiques (estime de soi, contrôle, sentiment de cohésion)
DG	Consommation de médicaments	RA	Limitation des activités
DH	Démographie et ménage	RP	Mouvements répétitifs
DV	Visites chez le dentiste	RS	Sécurité routière
ED	Niveau de scolarité	RT	Rationalité (échantillon supplémentaire du Manitoba, 1994-1995)
ES	Services d'urgence	SD	Renseignements socio-démographiques
EX	Examen de la vue	SH	Santé sexuelle

Fl	Équilibre et chutes (établissements)	SM	Usage du tabac
FS	Vaccination antigrippale	SP	Identificateurs d'échantillon (méthodologie)
GE	Identificateurs géographiques (méthodologie)	SS	Soutien social
GH	État de santé général	ST	Stress
HC	Utilisation des soins de santé	SV	Services de santé
HH	Ménage	TU	Bronzage et exposition aux rayons UV
HI	Renseignements sur la santé	TW	Incapacité au cours des deux dernières semaines
HS	État de santé	VS	Sécurité personnelle et violence
HV	VIH	WH	Santé de la femme : auto-examen des seins, examen des seins, mammographie et test de Papanicolaou
HW	Taille et poids	WT	Facteurs de pondération

12.2.3 Position 3 : Type d'enquête

A Enquête supplémentaire sur l'asthme
B Contenu de l'enquête auprès des échantillons supplémentaires des provinces - questions posées aux enfants
C Questions posées à l'échantillon principal qui seront répétées à chaque cycle
I Établissements de soins
K Questions posées aux enfants membres du panel
N Nord (Yukon / Territoires du Nord-Ouest)
P Contenu de l'enquête auprès des échantillons supplémentaires des provinces - questions posées aux adultes
S Enquête supplémentaire nationale (Enquête sur la promotion de la santé)
_ Questions particulières à un cycle, qui ne sont pas répétées chaque fois (stress en 1994-1995, accès aux services de santé en 1996-1997)
3 Variables d'administration de l'enquête pour la composante des ménages et la composante démographique (H03)
5 Variables d'administration de l'enquête pour la composante générale (H05)
6 Variables d'administration de l'enquête pour la composante santé (H06)

12.2.4 Position 4 : Variable de l'année de référence / du cycle

4	1994-1995	A	2004-2005
6	1996-1997	B	2006-2007
8	1998-1999	C	2008-2009
0	2000-2001	D	2010-2011
2	2002-2003	E	2012-2013
		F	2014-2015

12.2.5 Position 5 : Type de variable

_	Variable collectée	Variable qui figure directement sur le questionnaire
C	Variable codée	Variable codée à partir d'une ou de plusieurs variables collectées (p. ex., code de la classification type des industries (CTI))
D	Variable transversale dérivée	Variable calculée d'après une ou plusieurs variables collectées ou codées, ordinairement pendant le traitement au bureau central (p. ex., indice de l'état de santé).
F	Variable indicatrice	Variable calculée à partir d'une ou de plusieurs variables collectées (comme une variable dérivée), mais ordinairement par l'application informatique, aux fins de son utilisation ultérieure durant l'interview (p. ex., indicateur de travail).
G	Variable groupée	Variables collectées, codées, supprimées ou dérivées agrégées en un groupe (p. ex., groupes d'âge).
L	Variable longitudinale dérivée	Variable calculée d'après des variables provenant d'au moins deux cycles d'enquête.

12.2.6 Positions 6 à 8 : Nom de la variable

En général, les trois dernières positions correspondent au nom qui figure sur le questionnaire. On utilise des nombres dans la mesure du possible : Q1 devient 1. Dans le cas des questions de type «Cochez toutes les réponses pertinentes», on se sert des lettres correspondant à chaque catégorie possible de réponse : Q1 (cochez toutes les réponses pertinentes) devient 1A, 1B, 1C, etc. On représente les variables démographiques utilisées fréquemment par les analystes par trois lettres indicatrices plutôt que par un numéro de question; par exemple, «âge» est représenté par DHC6_AGE en 1996-1997. Dans les cas où les réponses à des groupes de questions portant sur le même sujet ont été collectées dans des sections du questionnaire portant des noms différents, on réserve la position 6 pour identifier la sous-section. Par exemple, la première question sur le stress chronique a été nommée ST_4_C1, la première question sur les facteurs de stress durant l'enfance et à l'âge adulte (traumatisme) est nommée ST_4_T1. Un autre exemple survient dans le cas des questions sur l'état de santé général de l'Enquête sur la promotion de la santé. Ces questions ont été réparties en trois sections aux fins de leur intégration au questionnaire et les noms des variables correspondantes reflètent cette situation, le caractère de la position 6 indiquant dans quelle section la question figure.

12.3 Téléaccès aux fichiers principaux

Pour pouvoir être considérés comme des fichiers à grande diffusion, les fichiers de microdonnées doivent répondre à des normes sévères de sécurité et de confidentialité, conformément à la *Loi sur la statistique,* afin de respecter le droit à la vie privée des répondants qui participent à l'enquête. Pour s'assurer du respect de ces normes, chaque fichier de microdonnées est soumis à un processus officiel d'examen destiné à confirmer qu'aucune personne ne pourra être identifiée. Les valeurs rares pour certaines variables susceptibles de permettre l'identification d'une personne sont supprimées du fichier ou agrégées en catégories moins détaillées, de façon à réduire au minimum le risque de divulgation de renseignements personnels. Fréquemment, ces variables sont les plus essentielles à l'analyse complète des données d'enquête. Puisqu'une quantité importante de ressources est investie dans la collecte de ces données, il est important de prendre des mesures pour tirer le plein potentiel analytique des fichiers de microdonnées afin de bien rentabiliser l'investissement statistique.

Le téléaccès aux fichiers principaux de l'enquête est un moyen de récolter ces fruits. On peut fournir à l'acheteur d'un produit de microdonnées un fichier principal «fictif» d'essai et le cliché d'enregistrement correspondant. Grâce à ces outils, il ou elle peut mettre au point son propre ensemble de programmes analytiques en se servant du fichier d'essai pour confirmer que les routines fonctionnent convenablement. Il ne lui reste plus qu'à

envoyer le code pour les totalisations personnalisées par Internet à nphs@statcan.ca. Le code est transmis au réseau interne protégé de Statistique Canada et traité en regard du fichier principal approprié de données de l'ENSP. Les résultats sont filtrés pour vérifier s'ils sont conformes aux normes de confidentialité et de fiabilité, puis, les données de sortie sont renvoyées au client. Ce service est gratuit.

Une deuxième méthode offerte à tous les clients consiste à demander au personnel spécialisé des Services personnalisés à la clientèle de la Division des statistiques sur la santé de produire des totalisations personnalisées. Ce service permet aux utilisateurs qui ne savent pas se servir de logiciels de totalisation d'avoir accès au fichier principal pour la préparation de leurs données personnalisées. Comme dans le cas du téléaccès, les résultats sont filtrés pour s'assurer qu'ils sont conformes aux normes de confidentialité et de fiabilité avant d'être diffusés. Contrairement au téléaccès, ce service est offert contre remboursement des frais.