

## TABLE DES MATIÈRES

	Page
1. Introduction.....	4
2. Contexte .....	6
3. Objectifs .....	7
4. Contenu de l'enquête.....	8
4.1 Critères de sélection.....	8
4.2 Modifications apportées au contenu du troisième cycle (1998-1999).....	9
4.3 Rétroaction des données et questions de suivi.....	10
4.4 Nouveau contenu du cycle 3 de l'ENSP (1998-1999).....	12
5. Plan d'échantillonnage .....	14
5.1 Plan d'échantillonnage pour 1998-1999 .....	14
5.1.1 Plan d'échantillonnage de la composante principale des ménages .....	14
5.1.2 Plan d'échantillonnage de 1998-1999 pour l'échantillon transversal supplémentaire.....	17
5.2 Plan d'échantillonnage pour 1994-1995 .....	18
5.2.1 Plan d'échantillonnage de la composante des ménages .....	19
5.2.2 Répartition de l'échantillon .....	19
5.2.3 Méthode d'exclusion .....	20
5.2.4 Sélection de l'échantillon .....	20
5.2.5 Plan d'échantillonnage au Québec .....	22
6. Collecte des données .....	24
6.1 Conception du questionnaire et méthode de collecte des données .....	24
6.2 Essais .....	24
6.3 Interviews .....	24
6.4 Non-réponse et dépistage.....	25
7. Traitement des données.....	27
7.1 Vérification .....	27
7.2 Codage .....	27
7.3 Création de variables dérivées et groupées.....	27
7.4 Facteurs de pondération .....	27
7.5 Suppression de renseignements confidentiels .....	28

8.	Qualité des données.....	29
8.1	Taux de réponse.....	29
8.1.1	Taux de réponse transversaux, Échantillon principal.....	29
8.1.2	Taux de réponse transversaux, Échantillon des logements non-répondants du cycle 1.....	31
8.1.3	Taux de réponse transversaux, Échantillon de remise à niveau tiré de l'EPA.....	32
8.1.4	Taux de réponse globaux.....	33
8.2	Erreurs relatives à l'enquête .....	34
9.	Lignes directrices pour la totalisation, l'analyse et la diffusion.....	37
9.1	Lignes directrices pour l'arrondissement.....	37
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation .....	38
9.2.1	Définitions : estimations de type nominal, estimations quantitatives .....	38
9.2.2	Totalisation d'estimations de type nominal.....	39
9.2.3	Totalisation d'estimations quantitatives .....	39
9.3	Lignes directrices pour l'analyse statistique.....	40
9.4	Lignes directrices pour la diffusion .....	41
10.	Tableaux de coefficients de variation approximatifs.....	42
10.1	Comment utiliser les tableaux de c.v. pour les estimations de type nominal ? .....	44
10.2	Exemples d'utilisation des tableaux de c.v. pour des estimations de type nominal ...	46
10.3	Utilisation des tableaux de c.v. pour calculer les intervalles de confiance.....	49
10.4	Exemple d'utilisation de tableaux de c.v. pour obtenir des limites de confiance.....	50
10.5	Utilisation des tableaux de c.v. pour effectuer un test Z .....	50
10.6	Exemple d'utilisation des tableaux de c.v. pour effectuer un test Z.....	51
10.7	Variances ou coefficients de variation réels .....	51
10.8	Seuils pour la diffusion des estimations relatives à l'ENSP.....	52
11.	Pondération .....	55
11.1	Pondération de l'échantillon transversal de l'ENSP de 1998-1990, Échantillon principal des ménages.....	56
11.1.1	Facteurs de pondération épurés .....	56
11.1.2	Corrections aux facteurs de pondération épurés.....	56
11.1.3	Corrections des facteurs de pondération appliqués aux membres des ménages .....	66
11.1.4	Corrections de la pondération pour les membres du panel.....	71
11.2	Pondération transversale de l'échantillon de l'ENSP de 1998-1999, Échantillon supplémentaire de remise à niveau .....	73
11.2.1	Facteurs de pondération de base de l'EPA .....	73
11.2.2	Sous-facteurs de pondération de l'EPA.....	74
11.2.3	Corrections supplémentaires des facteurs de pondération en fonction des sous-facteurs de pondération .....	74

11.3	Méthodes de pondération fondées sur celles de 1994-1995 pour les provinces sauf le Québec .....	75
11.3.1	Facteurs de pondération de base de l'EPA .....	75
11.3.2	Corrections supplémentaires des facteurs de pondération de base.....	76
11.4	Méthodes de pondération pour le Québec fondées sur celles de 1994-1995 .....	77
11.4.1	Facteurs de pondération de l'ESS.....	77
11.4.2	Facteur de pondération de base appliqué aux logements de l'ENSP .....	78
12.	Utilisation du fichier .....	79
12.1	Utilisation des facteurs de pondération.....	79
12.1.1	Facteur de pondération, Fichier général transversal, WT58.....	79
12.1.2	Facteur de pondération, Fichier sur la santé transversal, WT68 .....	79
12.2	Convention utilisée pour nommer les variables.....	79
12.2.1	Structure élémentaire des noms des variables .....	80
12.2.2	Positions 1 et 2 : Nom de la variable / section du questionnaire .....	81
12.2.3	Position 3 : Type d'enquête .....	82
12.2.4	Position 4 : Variable de l'année de référence du cycle.....	82
12.2.5	Position 5 : Type de variable .....	83
12.2.6	Positions 6 à 8 : Nom de la variable .....	83
12.3	Accès aux données des fichiers maîtres.....	84

## Liste des annexes

Annexe A : Questionnaire

Annexe B : Cliché d'enregistrement, Fichier de microdonnées Général

Annexe C : Cliché d'enregistrement, Fichier de microdonnées Santé

Annexe D : Dictionnaire des données, Fichier de microdonnées Général

Annexe E : Dictionnaire des données, Fichier de microdonnées Santé

Annexe F : Variables dérivées et groupées

Annexe G : Tableaux de C.V. fichier de microdonnées Général  
 Tableaux de C.V. Canada selon le groupe d'âge, fichier de microdonnées Général  
 Tableaux de C.V. selon les provinces et le Canada, fichier de microdonnées  
 Général

Annexe H : Tableaux de C.V. fichier de microdonnées Santé  
 Tableaux de C.V. Canada selon le groupe d'âge, fichier de microdonnées Santé  
 Tableaux de C.V. selon les provinces et le Canada, fichier de microdonnées Santé

## **1. Introduction**

L'Enquête nationale sur la santé de la population (ENSP) vise à recueillir des renseignements sur la santé de la population canadienne ainsi que des renseignements socio-démographiques connexes. Elle compte trois composantes : l'enquête auprès des ménages, l'enquête auprès des établissements de santé et l'enquête dans le Nord. Les présents fichiers de microdonnées à grande diffusion (FMGD) contiennent les données recueillies dans le cadre de la composante des ménages du troisième cycle de l'ENSP réalisé en 1998-1999.

La composante des ménages vise les membres 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. L'enquête a débuté en 1994 et aura lieu tous les deux ans pendant une période d'environ 20 ans. Trois cycles de collecte sont donc complétés pour chacune des composantes : l'ENSP Cycle 1(1994-1995), l'ENSP Cycle 2 (1996-1997) et l'ENSP Cycle 3 (1998-1999).

Pour le premier cycle, un échantillon d'environ 20,000 ménages a été sélectionné à partir de la base de sondage de l'Enquête sur la population active (EPA). Pour le troisième cycle, on s'est de nouveau servi de cette base de sondage pour sélectionner un échantillon supplémentaire de personnes ayant immigré récemment et de jeunes enfants, afin que l'échantillon soit représentatif de la population du Canada en 1998-1999.

L'ENSP recueille des renseignements généraux sur la santé de tous les membres des ménages sélectionnés pour participer à l'enquête et, dans chacun de ces ménages, une personne choisie au hasard lors du premier cycle fournit des renseignements plus détaillés sur sa santé. Au troisième cycle, environ 49 000 personnes ont répondu au questionnaire général et environ 17 000 ont répondu au questionnaire détaillé sur la santé.

Le questionnaire comprend, entre autres, des questions sur l'état de santé, l'utilisation des services de santé, les déterminants de la santé, les problèmes de santé chroniques et les limitations d'activité. L'utilisation des services de santé a été mesurée au moyen de questions sur les visites faites chez des professionnels de la santé, les soins hospitaliers et la consommation de médicaments. Les déterminants de la santé étudiés incluent l'usage du tabac, la consommation d'alcool et l'activité physique. Le nouveau contenu pour le troisième cycle de l'ENSP porte sur les antécédents médicaux de la famille, les soins personnels et la nutrition. On recueille également des renseignements tels que l'âge, le sexe, le niveau de scolarité, l'appartenance ethnique, le revenu du ménage et l'activité sur le marché du travail.

Le présent document a pour but de faciliter la manipulation de deux fichiers de microdonnées transversales contenant les résultats du cycle 3 de l'ENSP (1998-1999). Ces fichiers sont décrits de façon plus détaillée au chapitre 4 et dans les annexes.

Pour toute question au sujet des données ou de leur utilisation, veuillez vous adresser à :

- Soutien technique ou des renseignements généraux sur les données :  
Service d'aide aux utilisateurs de produits électroniques 1-800-949-9491
  
- Pour des totalisations personnalisées ou des renseignements généraux sur les données :  
Services personnalisés à la clientèle  
Division de la statistique de la santé 1-613-951-1746  
E-mail : [hd-ds@statcan.ca](mailto:hd-ds@statcan.ca)
  
- Pour des renseignements sur le téléaccès :  
Colette Koeune 1-613-951-1653  
Courriel : [nphs@statcan.ca](mailto:nphs@statcan.ca)  
Télécopieur : 1-613-951-4198
  
- Pour des renseignements sur le contenu de l'enquête :  
Mario Bédard 1-613-951-8933  
France Bilocq 1-613-951-6956  
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 les régimes de soins 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 broser 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 qui a été conçue en prenant pour critères la souplesse de l'enquête, ainsi que la validité, la fiabilité et l'actualité des données. Il fallait aussi que l'enquête puisse répondre à l'évolution des besoins, des politiques et des intérêts.

### **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 qui permettent de réaliser des études analytiques et qui aident à comprendre les déterminants de la santé ;
- Recueillir des données sur les corrélations entre la santé et les facteurs économiques, sociaux, démographiques, professionnels et environnementaux ;
- Permettre de mieux comprendre 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 coupler les données de l'enquête à 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 l'orientation générale de l'ENSP, notamment en ce qui concerne le type de renseignements à recueillir. La première section de ce chapitre présente les critères généraux utilisés pour choisir le contenu de l'enquête et résume le contenu des diverses sections. La seconde section décrit en détail les modifications apportées au contenu de l'enquête pour réaliser le troisième cycle de l'ENSP. La section suivante met l'accent sur les variables des cycles précédents qui ont été utilisées lors de la collecte du troisième cycle. Enfin, la dernière section décrit le nouveau contenu de l'enquête de 1998-1999.

### **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 mieux faire 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 aux régimes de soins 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, on a évalué l'état de santé à l'aide de questions sur la perception personnelle de la santé, la capacité fonctionnelle, les problèmes de santé chroniques et 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 déterminants de la santé avaient trait à l'usage du tabac, la consommation d'alcool et l'activité physique. En outre, le contenu thématique du troisième cycle de l'enquête englobait les soins personnels, les antécédents médicaux de la famille et la nutrition. On a de nouveau posé les questions sur le sentiment de cohésion qui avaient été posées lors du premier cycle. Les questions sur les caractéristiques démographiques et socio-économiques couvraient 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 promotion santé (EPS) de 1998-1999 à l'ENSP au lieu de l'effectuer séparément à titre d'enquête complémentaire comme en 1994-1995. Des questions supplémentaires portaient sur les variantes de l'usage du tabac, c'est-à-dire son usage sous d'autres formes que la cigarette.

L'annexe A contient le questionnaire complet de l'ENSP de 1998-1999.

#### **4.2 Modifications apportées au contenu du troisième cycle (1998-1999)**

L'intégration du contenu visant les enfants de moins de 12 ans dans le cheminement ordinaire du questionnaire constitue une amélioration importante par rapport au cycle précédent.

##### **Composante Générale - Modifications par section**

###### **Limitation des activités**

RESTR-Q1D      Ajout de «ou l'école».

###### **Renseignements socio-démographiques**

SOCIO-Q6      Ajout d'une instruction à l'intention de l'intervieweur concernant les bébés.

###### **Niveau de scolarité**

EDUC-Q7      Ajout du mot «jamais» afin que le répondant comprenne qu'il s'agit du plus haut niveau cumulatif de scolarité atteint.

###### **Revenu**

INC-Q1      L'instruction «passer à la section suivante» indiquée pour la réponse «aucune» à la question Q1 a été remplacée par «passer à INCOM-Q3».

INC-Q3      On a d'abord posé la question sur le revenu exact du ménage. La question en cascade n'a été posée qu'en cas de non-réponse.

##### **Composante Santé - Modifications par section**

Le contenu du supplément effectué pour l'Alberta en 1996 a été supprimé. Le contenu thématique sur l'accès aux services de santé a été supprimé, sauf les

questions de base sur la mammographie, le test de Papanicolaou (test PAP) et la tension artérielle.

#### Mesures de santé préventives

Les questions sur la tension artérielle, la mammographie et le test PAP ont été regroupées sous cette rubrique, comme en 1994. Ces questions n'ont été séparées qu'en 1996 afin de les poser avec les questions thématiques sur l'accès aux services. Toutes les questions sur la santé féminine ont été regroupées dans cette section, y compris celles sur la grossesse et la naissance du dernier enfant.

#### Assurance

Toutes les questions concernant la couverture par une assurance ont été regroupées dans cette section.

#### Blessures

##### IN-Q7

En 1994, cette question a été posée aux personnes qui travaillaient au moment de l'interview. Cette démarche était incorrecte car une personne pouvait ne pas travailler au moment de l'interview à cause d'une blessure. En 1998, la question a été posée à toute personne de 15 ans et plus, quelle que soit sa situation d'activité.

#### Consommation de médicaments

##### DRG-Q3

Ajout d'une question de suivi pour chaque médicament pris, afin de déterminer s'il s'agit ou non d'un médicament prescrit.

#### Soutien social

Remplacement de toute la section par une nouvelle mesure du soutien social (Étude sur les résultats médicaux (ERM)). Consulter la documentation sur les variables dérivées.

#### Consentement au partage de l'information

Modification de l'énoncé de la question sur le consentement au partage de l'information afin de tenir compte de tous les cycles de l'enquête. L'expression «les données recueillies grâce à ce sondage» a été remplacée par «l'information recueillie lors des interviews effectuées dans le cadre de cette enquête».

### **4.3 Rétroaction des données et questions de suivi**

Afin de réduire le fardeau de réponse, les questions posées lors des cycles précédents et dont la réponse n'a pas changé dans le temps (le pays de naissance par exemple) n'ont pas été posées lors du cycle 3 de l'ENSP. Les variables pour lesquelles la réponse peut changer au cours du temps, mais uniquement dans certaines conditions (p. ex., niveau de scolarité) ont été mise à jour au besoin. Certaines réponses recueillies lors des cycles précédents ont été rappelées au répondant lors de l'interview du troisième cycle. Cette technique de rappel a permis d'améliorer la qualité des estimations.

#### Limitation des activités

On s'est servi des données indiquant si le répondant souffrait ou non d'une incapacité en 1996. En cas de changement de situation, on a posé des questions pour obtenir une explication.

#### Problèmes de santé chroniques

Pour chaque répondant, on a utilisé les réponses du cycle 2 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 cérébro-vasculaire) 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 utilisé des indicateurs montrant que les renseignements concernant le pays de naissance et l'origine ethnique ont été recueillis. 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 utilisé 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, le secteur d'activité et le genre de tâches correspondant à l'emploi principal mentionné à l'interview de 1996-1997. Aux répondants qui ont dit avoir travaillé l'année précédente, on a demandé de confirmer le nom de l'employeur. Dans le cas de changement, les renseignements ont été saisis.

#### Hypertension, mammographie, test PAP, usage du tabac

Lors du premier cycle et du deuxième cycle, on a demandé à la personne interrogée si on avait «déjà» pris sa tension artérielle (ou si elle avait «déjà» passé une mammographie, etc.). Dans certains cas, les données recueillies lors des deux cycles précédents étaient contradictoires. En 1998, on a répété les questions. Dans les cas où la personne a indiqué qu'elle n'avait pas subi le test en question mais qu'elle avait dit l'avoir subi lors d'un cycle antérieur, on a posé des questions supplémentaires.

#### Numéro d'assurance maladie

On a utilisé un indicateur montrant si le numéro d'assurance maladie recueilli lors d'un cycle antérieur était «valide». Si le numéro d'assurance maladie était invalide et que le répondant a donné le même numéro que celui enregistré lors du dernier cycle, on a demandé au répondant de redonner son numéro d'assurance maladie.

#### **4.4 Nouveau contenu du cycle 3 de l'ENSP (1998-1999)**

##### **Composante Générale**

###### Utilisation des soins de santé

On a ajouté une question à réponse écrite pour déterminer la raison pour laquelle des soins ont été obtenus aux États-Unis.

###### Changement de résidence

Pour des raisons de remise à niveau de l'échantillon, on a ajouté cette section afin de déterminer le lieu de résidence des nouveaux membres des ménages qui font partie du panel longitudinal depuis le 1<sup>er</sup> octobre 1994.

###### Revenu personnel

Dans la composante générale, on a demandé aux répondants âgés de 15 ans et plus d'indiquer leur revenu personnel exact. Cette question a été suivie de questions en cascade en cas de non-réponse à la question sur le revenu exact.

###### Précarité alimentaire

Trois questions filtres ont été posées pour repérer les ménages «courant le risque » d'être en situation de précarité alimentaire.

##### **Composante Santé**

###### Hystérectomie

De nouvelles questions ont été posées en vue de déterminer la prévalence de l'hystérectomie, l'âge et les raisons pour laquelle elle a été subie.

###### Soins personnels

Ces questions font partie du contenu thématique de 1998-1999.

###### Antécédents médicaux de la famille

Ces questions font partie du contenu thématique de 1998-1999.

###### Nutrition

On a ajouté ces questions afin d'intégrer certaines questions de référence sur la nutrition. Ce sujet a été proposé comme contenu thématique pour 1998-1999. Un essai pilote avec interview papier et crayon a été réalisé à l'automne de 1997. Les questions sur la nutrition, dont la mise au point se poursuit, seront peut-être incluses comme contenu thématique lors d'un cycle futur.

###### Sentiment de cohésion

Ce groupe de questions a été posé lors du premier cycle en 1994. Il était prévu de poser ce groupe de questions de façon périodique car, selon certains, le sentiment de cohésion d'une personne de plus de 35 ans demeure relativement stable. Les nouvelles données permettront de vérifier la validité de cette hypothèse et seront, de

plus recueillies pour la première fois pour l'échantillon de remise à niveau et pour les personnes qui avaient moins de 18 ans en 1994.

Usage du tabac

SMK-Q5A      Question sur la quantité totale posée aux anciens fumeurs.

SMK-Q5B      Question sur le nombre de cigarettes et la fréquence de la consommation de cigarettes posée aux fumeurs occasionnels.

Variantes de l'usage du tabac

Ces questions de l'Enquête promotion santé serviront de questions de référence sur l'usage du cigare, de la pipe, du tabac à priser et du tabac à chiquer.

Certaines questions de l'Enquête promotion santé ont été retenues à titre de questions de base et seront désormais posées lors de chaque cycle.

## **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 1998-1999**

#### **5.1.1 Plan d'échantillonnage de la composante principale des ménages**

Au cours du premier cycle de l'ENSP, 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 troisième cycle, une distinction est faite entre l'échantillon choisi à des fins longitudinales et celui choisi à des fins transversales.

L'échantillon longitudinal de 1998-1999 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 ont été interviewées dans le cadre de l'Enquête nationale longitudinale sur les enfants et les jeunes (ELNEJ) et qui ont été ajoutées à 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 aux cycles précédents et 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 a été élargi afin de produire des estimations transversales pour 1998-1999 (voir 5.1.2). L'échantillon supplémentaire a été ajouté afin de compenser pour l'attrition et pour tenir compte des personnes initialement absentes, c'est-à-dire les enfants en bas âge et les nouveaux immigrants. Dans l'ensemble, l'échantillon transversal du cycle 3 est légèrement plus grand que celui du cycle 1.

**Taille de l'échantillon longitudinal principal selon la province courante**

Nombre de personnes

<b>Province</b>	<b>Échantillon principal de 1994-1995<sup>1</sup></b>	<b>Échantillon Principal de 1998-1999</b>	<b>% de suivi de l'échantillon principal de 1994-1995</b>
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

Puisque aucune nouvelle unité longitudinale n'a été sélectionnée au cours des deuxième et troisième cycles, la population comprise dans l'échantillon longitudinal en 1998-1999 est âgée de quatre ans ou plus. Cela veut dire qu'aucune personne sélectionnée n'était âgée entre 0 et 3 ans au moment du troisième 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 troisième cycle a aussi eu pour effet que les personnes qui ont immigré au Canada au cours des quatre 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, étant donné qu'un échantillon supplémentaire a été tiré pour ces sous-populations, des estimations peuvent être faites pour la population totale transversale de 1998-1999. Tous les enfants âgés entre 0 et 3 ans qui vivent en ménage avec les répondants longitudinaux, de même que les personnes ayant immigré récemment et qui ont joint ces ménages, sont aussi représentés par l'échantillon principal.

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

<sup>1</sup> Échantillon réel excluant les logements non admissibles, vacants, en construction ou rejetés à la suite de l'application de la méthode d'exclusion.

comprend tous les membres des ménages qui vivent avec des répondants longitudinaux. La population cible de l'échantillon supplémentaire 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.

<b>Situation du répondant longitudinal</b>	<b>Mesures prises</b>
Décédé	Dans le cas des répondants longitudinaux dont on a déterminé qu'ils étaient décédés au moment de l'interview du troisième 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 autres membres du ménage en se servant du questionnaire visant les ménages. Ces répondants font partie du fichier 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 1998-1999 pour l'échantillon transversal supplémentaire**

Deux petits échantillons supplémentaires (ou aussi appelés échantillons de remise à niveau) ont été tirés afin de rendre l'échantillon transversal plus représentatif de la population de 1998-1999. Le premier a été créé pour tenir compte de la population initialement absente, c'est-à-dire les personnes n'ayant pas pu être échantillonnées en 1994-1995. Le second a été créé pour compenser pour l'attrition de l'échantillon depuis le cycle 1. De la même façon que pour l'échantillon principal, la composante générale du questionnaire a été administrée à tous les membres des ménages répondants. Un membre du ménage a aussi été sélectionné pour répondre à la composante santé. Ces échantillons supplémentaires ont été combinés à l'échantillon principal pour produire un seul gros fichier de données transversales.

Le premier échantillon supplémentaire était un échantillon aléatoire de la population initialement absente, enfants nés en 1995 ou après et immigrants entrés au Canada depuis le début de 1995. Quatre différents groupes de rotation de l'Enquête sur la population active (EPA) ont été utilisés comme base pour ces populations soit un groupe pour chacune des périodes de collecte de l'ENSP. Lorsque chaque groupe de rotation atteignait son dernier mois dans l'échantillon de l'EPA, un supplément demandant le pays de naissance de chaque membre du ménage était ajouté au questionnaire de l'EPA. En utilisant les réponses à cette question de même que la liste des membres du ménage, tous les nouveaux immigrants et les enfants en bas âge ont été identifiés, le premier janvier 1995 étant la date limite utilisée.

Tous les 422 ménages d'immigrants ont été interviewés par l'ENSP, trois ou quatre mois après que leurs groupes de rotation correspondants aient été libérés par l'EPA. Les enfants en bas âge ont été échantillonnés à partir de seulement deux des groupes de rotation et interviewés seulement aux trimestres 2 et 3 à cause de conflits d'horaire avec l'ELNEJ. Un total de 758 enfants en bas âge ont été interviewés, ce qui est approximativement la moitié de tous ceux disponibles dans les deux groupes de rotation. Au trimestre 3, l'ELNEJ avait besoin d'un petit échantillon d'enfants de moins d'un an. Une portion de ces enfants ont donc été interviewés par l'ELNEJ et non par l'ENSP. Une procédure spéciale de stabilisation a été réalisée lors de la pondération pour tenir compte de ce cas spécial et aussi des ménages ayant plusieurs enfants.

Puisque la liste des membres du ménage était déjà connue suite à l'interview de l'EPA, la personne ayant à répondre au questionnaire santé a été choisie avant que les dossiers soient envoyés sur le terrain pour la collecte et le dépistage était fait si la personne sélectionnée avait déménagé dans les trois ou quatre mois suivants l'interview pour l'EPA. Pour les ménages contenant des immigrants, on a planifié de faire les interviews en personne afin de faciliter le dépistage,

alors que les interviews pour les enfants en bas âge étaient menées au téléphone, pour réduire les coûts.

Le second échantillon supplémentaire a été tiré pour remplacer la partie d'échantillon perdue suite à l'attrition depuis le cycle 1. Les individus vivant dans les logements qui faisaient partie de la base de sondage originale mais dont le ménage n'avait pas répondu en 1994-1995 remplissaient les conditions pour être dans l'échantillon supplémentaire. Un total de 2 598 ménages ont été contactés et invités à participer. Puisque seulement l'adresse était connue pour ces logements, il a été planifié de mener toutes les interviews par des visites sur place. Lors du contact initial, les intervieweurs dressaient la liste des membres du ménage. À cause d'un problème avec l'application informatique de collecte de données (c'est-à-dire la version informatisée du questionnaire), aucun enfant âgé de 0 à 11 ans n'a été sélectionné pour recevoir le questionnaire santé pour l'échantillon supplémentaire, ce qui a été pris en considération dans les procédures de pondération.

**Taille de l'échantillon transversal de 1998-1999 selon la province courante**  
Nombre de ménages

Province	Échantillon supplémentaire	Échantillon principal	Total
Terre-Neuve	140	1 082	1 222
Île-du-Prince-Édouard	167	1 037	1 204
Nouvelle-Écosse	228	1 085	1 313
Nouveau-Brunswick	191	1 125	1 316
Québec	590	3 000	3 590
Ontario	1 392	4 307	5 699
Manitoba	205	1 205	1 410
Saskatchewan	213	1 168	1 381
Alberta	221	1 544	1 765
Colombie-Britannique	431	1 723	2 154
Canada	3 778	17 276	21 054

## **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)<sup>2</sup>. Selon cette méthode, l'échantillon a été réparti proportionnellement à  $\sqrt{(0.804Wh^2 + 1/12^2)}$ , où  $W_h$  représente la proportion des ménages dans la province / le territoire ( $h=1, \dots, 12$ ) 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 initialement réparti de façon proportionnelle à la taille de la population. Les unités d'échantillonnage additionnelles achetées par certaines provinces et l'utilisation d'une méthode d'exclusion, ci-après décrite, ont eu un effet sur la répartition infra provinciale 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). La taille des échantillons a été augmentée en fonction du nombre de ménages que l'on s'attendait à rejeter par la méthode d'exclusion au moment de la sélection. La

---

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

taille d'échantillon réelle, mentionnée plus tôt dans ce chapitre, exclut les ménages rejetés par la méthode d'exclusion ainsi que les logements hors champs d'enquête, comme les logements en construction.

### **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 sur-représentées.

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 Sélection 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 des 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 groupes de rotation par province, un groupe de rotation 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 groupes de rotation mais que la taille de l'échantillon requise représente seulement 1,5 groupes de rotation.

En raison des besoins particuliers de l'ENSP, deux modifications ont été apportées à cette stratégie d'échantillonnage. Le nombre de groupe rotation «réservés» a été déterminé au niveau des strates plutôt qu'au niveau provincial, afin de répondre aux exigences relatives à la répartition infra provinciale 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.

À la 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 dix logements par grappe dans les grandes villes et de vingt ou trente 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 ménages comptant des jeunes âgés de moins de 25 ans et les ménages contenant plus d'un membre, mais sans enfant ni jeune). Le type de ménages dans le cas des non-répondants de l'ESS a été déterminé par le personnel de l'ENSP.

La taille de l'échantillon de l'ENSP a d'abord été répartie parmi les quatre catégories de densité urbaine. Afin d'éviter la concentration de l'échantillon à Montréal, la répartition a été proportionnelle à  $\sqrt{(2Wh^2 + 1/4^2)}$  où  $W_h$  représente la part de la population pour la catégorie  $h=1,2,3,4$ . 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 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 presque été 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 «adulte» et un échantillon «enfant», ces termes ayant la même signification que dans les autres provinces. Les ménages de l'échantillon «enfants» des trimestres 1 et 2 ont été répartis 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 «adulte» et un échantillon «enfant», 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.

## **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 saisie 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 ou durant une interview précédente.

### **6.2 Essais**

Deux essais sur le terrain ont été effectués. Quatre des bureaux régionaux de Statistique Canada ont participé à ces essais. Les interviews ont été menées par des intervieweurs expérimentés de l'Enquête sur la population active. Ces deux essais visaient principalement à observer les réactions des répondants, à estimer la durée des diverses sections du questionnaire, à déterminer les taux de réponse et à évaluer les questions avec rétroaction. Les opérations et les procédures sur le terrain, la formation des intervieweurs et l'application informatique de collecte de donnée ont aussi été soumis à des essais.

En plus des essais sur le terrain, l'application informatique de collecte des données a été soumise 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 de l'application s'est poursuivie de façon continue jusqu'au début de l'enquête principale.

### **6.3 Interviews**

Les intervieweurs ont été formés spécifiquement pour mener des enquêtes selon la méthode d'interview assistée par ordinateur. La collecte des données a été répartie en quatre trimestres (juin, août et novembre 1998 et février 1999). Une autre collecte s'est tenue en juin 1999 avec des tentatives supplémentaires de dépistage des non-



répondants des trimestres précédents. On a appliqué à la collecte des données de l'ENSP la structure de supervision et de vérification de l'EPA.

On a d'abord communiqué par téléphone avec les répondants de l'échantillon principal et de l'échantillon de remise à niveau de ménages comptant de jeunes enfants. 95 % des interviews ont été effectuées par téléphone. On a procédé à des interviews sur place si les répondants n'avaient pas le téléphone, dans le cadre d'un dépistage en personne d'un répondant ou à la demande d'un répondant. L'intervieweur a également rendu visite en personne aux répondants de l'échantillon de remise à niveau sélectionnés parmi les non-répondants du premier cycle et de l'échantillon de remise à niveau de ménages formés de nouveaux immigrants. L'interview complète auprès de chaque ménage a duré en moyenne une heure.

Dans chaque logement, on a recueilli des renseignements généraux sur tous les membres du ménage auprès de la personne qui se trouvait à la maison au moment de l'appel de l'intervieweur. Cependant, en 1998 on a encouragé les intervieweurs à commencer par les répondants choisis pour faire partie du panel longitudinal. Environ 15 % des renseignements recueillis pour cette partie de l'interview (membres du panel longitudinal de 12 ans et plus) l'ont été par personne interposée.

Dans le cas des répondants longitudinaux, la déclaration par personne interposée pour la Composante santé n'a été permise que pour des raisons de maladie ou d'incapacité. Ces déclarations par personne interposée représentent 2,7 % des données recueillies pour les répondants de 12 ans et plus. En revanche, toutes les interviews auprès des répondants longitudinaux âgés de moins de 12 ans ont été effectuées par personne interposée.

#### **6.4 Non-réponse et dépistage**

On a demandé aux intervieweurs de déployer tous les efforts raisonnables pour que les membres des ménages admissibles à l'enquête acceptent de répondre au questionnaire de l'ENSP. Aux personnes qui ont refusé au départ de participer à l'ENSP, le Bureau Régional concerné a envoyé une lettre pour souligner l'importance de l'enquête et de la coopération du ménage. Puis, l'intervieweur a fait un deuxième appel (ou une deuxième visite). Si le moment de cet appel (ou de cette visite) était mal choisi, l'intervieweur a convenu avec la personne intéressée d'un moment plus opportun pour rappeler ou pour revenir. Dans les cas où il n'y avait personne au domicile, plusieurs tentatives ont été faites pour contacter le ménage.

De nombreuses stratégies ont été adoptées pour réduire le nombre de cas de non-réponse. Avant le début des interviews, on a déterminé d'après les résultats d'essais la taille maximale recommandée de la tâche attribuée à chaque intervieweur, afin de permettre un suivi efficace des cas de non-contact (autrement dit, pour éviter de surcharger les intervieweurs). La formation offerte aux intervieweurs incluait divers moyens de réduire le nombre de cas de non-contact (par exemple, faire des visites ou des appels à divers moments de la journée) en se servant des renseignements de base recueillis au cours de l'interview précédente.

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.

Le fait de ne pas dépister un répondant longitudinal a été considéré comme un autre type de non-réponse. Les intervieweurs disposaient de plusieurs moyens 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, comme les annuaires téléphoniques locaux et l'assistance annuaire. Les cas pour lesquels ces démarches n'ont donné aucun résultat ont été transmis à un intervieweur chevronné spécialement formé pour dépister les 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 cumulatif lié à l'impossibilité de dépister le répondant longitudinal se chiffre à 2,1 %, chiffre extrêmement faible.

## **7. Traitement des données**

### **7.1 Vérification**

La vérification des données s'effectue au moment de l'interview assistée par ordinateur (IAO). L'intervieweur ne peut saisir des valeurs situées hors de la fourchette permise et les erreurs d'enchaînement sont contrôlées à l'aide du système programmé de sauts de questions (commandes «passer à») de l'IAO. 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 sont donnés, mais aucune mesure corrective n'est prise (par exemple, lorsqu'un intervieweur saisit des réponses contradictoires). Comme aucune mesure corrective n'est prise en pareil cas, on a élaboré des vérifications qui sont exécutées au Bureau central après la collecte des données. En général, les incohérences sont corrigées en sélectionnant la valeur «non déclaré» pour l'une des variables en cause ou les deux. Aucune imputation n'est effectuée.

### **7.2 Codage**

Pour plusieurs questions auxquelles il est possible d'inscrire une réponse, on a codé les renseignements inscrits de façon à les classer dans de nouvelles catégories distinctes ou dans une catégorie énumérée existante si les renseignements inscrits étaient le calque d'une catégorie déjà énumérée. Dans la mesure du possible (par exemple pour la profession, l'industrie, les maladies), on a codé les renseignements conformément aux systèmes de classification type utilisés pour le Recensement de la population ou pour d'autres enquêtes de Statistique Canada, comme 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 des fichiers de microdonnées d'après certains éléments 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 les fichiers 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 SMC8\_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 les poids d'échantillonnage est décrite en détail au chapitre 11.

## **7.5 Suppression de renseignements confidentiels**

Il convient de mentionner que les fichiers de microdonnées «à grande diffusion» diffèrent à bien des égards des fichiers «maîtres» de l'enquête conservés 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 l'établissement d'une valeur plafond des variables. Statistique Canada offre différentes options permettant le calcul de statistiques incluant les renseignements exclus des fichiers de microdonnées. Ces options sont décrites dans la section 12.3. Les estimations produites sont transmises 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 de l'ENSP doit tenir compte de l'augmentation de l'échantillon par l'ajout des logements non-répondants du premier cycle et par l'échantillon de remise à niveau tiré de l'EPA (composé d'immigrants et d'enfants). Ces échantillons de remise à niveau ne sont utilisés que dans le but de produire des estimations transversales. Par conséquent, les taux de réponse sont calculés séparément pour l'échantillon principal, pour l'échantillon des logements non-répondants du premier cycle et pour l'échantillon de remise à niveau tiré de l'EPA, puis ils sont calculés pour l'échantillon global. Les taux de réponse calculés pour 1998-1999 sont résumés dans le tableau ci-après.

**Tableau des taux de réponse pour 1998-1999**

Niveau	Échantillon principal	Échantillon des logements non-répondants du cycle 1	Échantillon de remise à niveau tiré de l'EPA	Échantillon global
Ménages	89,7%	71,5%	92,5%	87,6%
Personnes sélectionnées	98,8%	96,4%	98,8%	98,5%

Suit la description de la méthode de calcul du taux de réponse des ménages et du taux de réponse des personnes sélectionnées appliquée au fichier transversal. On notera que, dans les quatre sections qui suivent, on calcule le taux de réponse des personnes sélectionnées d'après le nombre de ménages répondants. Pour avoir une idée du taux global de réponse des personnes sélectionnées, il suffit de multiplier les deux taux.

#### 8.1.1 Taux de réponse transversaux, Échantillon principal

##### *Taux de réponse des ménages (M)*

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

On calcule le taux de réponse transversal de 1998-1999 pour l'échantillon *principal* en se fondant sur l'ensemble des ménages qui continuent à participer à l'enquête, sauf ceux hors du champ de l'enquête (par exemple, une personne sélectionnée 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 (H05). Le taux de réponse transversal de l'ENSP, au niveau national, se chiffre à **89,7%**. Au niveau provincial, ce taux varie de 85,3% en Colombie-Britannique à 93,1% à Terre-Neuve.

***Taux de réponse des personnes sélectionnées (PS)***

On peut représenter le taux de réponse des personnes comme le nombre de composantes santé (H06) qui ont *effectivement* été remplies comparativement au nombre de composantes santé qui *auraient dû* être remplies.

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

Pour l'échantillon principal, le taux de réponse des personnes sélectionnées de l'ENSP se chiffre à **98,8%** au niveau national et varie de 98,2% au Québec à 99,6% en Saskatchewan.

***Renseignements pertinents pour le calcul des taux de réponse :***

Nombre de ménages qui continuent à participer à l'enquête : 17 276

Nombre de ménages hors du champ de l'enquête : 880

Nombre de répondants au niveau des ménages : 14 702

Nombre de répondants au niveau des personnes sélectionnées : 14 520

Nombre de non-répondants au niveau des ménages : 1 694

Nombre de non-répondants au niveau des personnes sélectionnées : 182

Calcul du taux de réponse des ménages :

$$\text{Taux M} = \frac{14\,702}{17\,276 - 880} = \frac{14\,702}{16\,396} = 89,7\%$$

Calcul du taux de réponse des personnes sélectionnées :

$$\text{Taux PS} = \frac{14\,520}{14\,702} = 98,8\%$$

### 8.1.2 Taux de réponse transversaux, Échantillon des logements non-répondants du cycle 1

#### *Taux de réponse des ménages (M)*

# de ménages répondants

---

# de ménages vivants dans les log. non - répondants du cycle 1 faisant partie du champ de l'enquête

On calcule le taux de réponse transversal de 1998-1999 de l'échantillon des ménages vivants dans les logements non-répondants du cycle 1 en se fondant sur un échantillon tiré parmi les logements non-répondants de l'échantillon principal du premier cycle, à l'exception de quelques logements en particulier qui sont devenus hors du champ de l'enquête (code de statut= 013, 017, 023, 024)<sup>3</sup>. Un ménage répondant est un ménage dont *au moins* un membre a répondu à la composante générale du questionnaire (H05). Le taux de réponse transversal de l'ENSP, au niveau national, se chiffre à **71,5%**. Au niveau provincial, ce taux varie de 59,5% à l'Île-du-Prince-Édouard à 76,5% en Ontario.

#### **Taux de réponse des personnes sélectionnées (PS)**

On peut représenter le taux de réponse des personnes sélectionnées comme le nombre de composantes santé (H06) qui ont *effectivement* été remplies comparativement au nombre de composantes santé qui *auraient dû* être remplies.

# de personnes qui ont répondu au questionnaire H06

---

# de ménages répondants (c.-à-d. remplissant les conditions requises pour répondre)

Pour la portion des logements non-répondants du cycle 1, le taux de réponse des personnes sélectionnées pour répondre à la composante santé de l'ENSP se chiffre à **96,4%** au niveau national et varie de 94,4% en Colombie-Britannique à 100,0% à Terre-Neuve.

#### **Renseignements pertinents pour le calcul des taux de réponse :**

Nombre de logements non-répondants de l'échantillon principal  
du premier cycle :

2 432

---

<sup>3</sup>013=Entreprise, établissement ou autre non-résidentiel.  
017=Autre logement non admissible (ex. ambassade).  
023=Logement en construction ou démoli.  
024=Logement vacant.

Nombre de ménages hors du champ de l'enquête : 8

Nombre de répondants au niveau des ménages : 1 732

Nombre de répondants au niveau des personnes sélectionnées : 1 670

Nombre de non-répondants au niveau des ménages : 692

Nombre de non-répondants au niveau des personnes sélectionnées : 62

Calcul du taux de réponse des ménages :

$$\text{Taux M} = \frac{1\,732}{2\,432 - 8} = \frac{1\,732}{2\,424} = 71,5\%$$

Calcul du taux de réponse des personnes sélectionnées :

$$\text{Taux PS} = \frac{1\,670}{1\,732} = 96,4\%.$$

### **8.1.3 Taux de réponse transversaux, Échantillon de remise à niveau tiré de l'EPA**

#### **Taux de réponse des ménages (M)**

$$\frac{\text{\# de ménages répondants}}{\text{\# de ménages sélectionnés à partir de l'EPA faisant partie du champ de l'enquête}}$$

On calcule le taux de réponse transversal de 1998-1999 de l'échantillon de remise à niveau tiré de l'EPA en se fondant sur l'ensemble des ménages qui ont été sélectionnés à partir de l'EPA. Un ménage répondant est un ménage dont *au moins* un membre a répondu à la composante générale du questionnaire (H05). Le taux de réponse transversal de l'ENSP, au niveau national, se chiffre à **92,5%** (92,5% pour les enfants et **92,7%** pour les nouveaux immigrants). Au niveau provincial, ce taux varie de 89,6% au Nouveau-Brunswick à 97,3% en Saskatchewan.

#### **Taux de réponse des personnes sélectionnées (PS)**

On peut se représenter le taux de réponse des personnes sélectionnées comme le nombre de composantes santé (H06) qui ont *effectivement* été remplies comparativement au nombre de composantes santé qui *auraient dû* être remplies.

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



Pour l'échantillon de remise à niveau tiré de l'EPA, le taux de réponse des personnes sélectionnées pour répondre à la composante santé de l'ENSP se chiffre à **98,8%** au niveau national (**100,0%** pour les enfants et **96,5%** pour les nouveaux immigrants) et varie de 97,6% en Alberta à 100,0% à Terre-Neuve, à l'Île-du-Prince-Édouard, en Nouvelle-Écosse, au Nouveau-Brunswick, au Québec et en Saskatchewan.

***Renseignements pertinents pour le calcul des taux de réponse :***

Nombre de ménages sélectionnés à partir de la base de sondage de l'EPA :	1 179
- Enfants :	758
- Nouveaux immigrants :	421
Nombre de ménages hors du champ de l'enquête :	26
Nombre de répondants au niveau des ménages :	1 067
Nombre de répondants au niveau des personnes sélectionnées :	1 054
Nombre de non-répondants au niveau des ménages :	86
Nombre de non-répondants au niveau des personnes sélectionnées :	13

Calcul du taux de réponse des ménages :

$$\text{Taux M} = \frac{1\,067}{1\,179 - 26} = \frac{1\,067}{1\,153} = 92,5\%$$

Calcul du taux de réponse des personnes sélectionnées :

$$\text{Taux PS} = \frac{1\,054}{1\,067} = 98,8\%$$

#### **8.1.4 Taux de réponse globaux**

##### **Taux de réponse des ménages (M)**

*# de ménages répondants*

---

*tous les ménages sélectionnés pour le cycle 3 de l'ENSP faisant partie du champ de l'enquête*

Pour l'échantillon principal, l'échantillon des logements non-répondants du cycle 1 et pour l'échantillon de remise à niveau tiré de l'EPA confondus, le taux de réponse transversal se chiffre à **87,6%** au niveau national. Au niveau provincial, le taux varie de 83,6% en Colombie-Britannique à 91,0% à Terre-Neuve.

### **Taux de réponse des personnes sélectionnées (PS)**

On peut se représenter le taux de réponse des personnes sélectionnées comme le nombre de composantes santé (H06) qui ont *effectivement* été remplies comparativement au nombre de composantes santé qui *auraient dû* être remplies.

*# de personnes qui ont répondu au questionnaire H06*

*# de ménages répondants (c.-à-d. remplissant les conditions requises pour répondre)*

Le taux de réponse des personnes sélectionnées est égal à **98,5%** au niveau national et varie de 98,2% en Nouvelle-Écosse et au Québec à 99,3% en Saskatchewan.

#### ***Renseignements pertinents pour le calcul des taux de réponse :***

Nombre de ménages sélectionnés pour le cycle 3 de l'ENSP : 20 887

Nombre de ménages hors du champ de l'enquête : 914

Nombre de répondants au niveau des ménages : 17 501

Nombre de répondants au niveau des personnes sélectionnées : 17 244

Nombre de non-répondants au niveau des ménages : 2 472

Nombre de non-répondants au niveau des personnes sélectionnées : 257

Calcul du taux de réponse des ménages :

$$\text{Taux M} = \frac{17\,501}{20\,887 - 914} = \frac{17\,501}{19\,973} = 87,6\%$$

Calcul du taux de réponse des personnes sélectionnées :

$$\text{Taux PS} = \frac{17\,244}{17\,501} = 98,5\%$$

## **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 ces fichiers 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 :

$$\left( \frac{.003}{.24} \right) \times 100\% = 1.25\%$$

## **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 leur permettre de reproduire les chiffres publiés par Statistique Canada et de produire aussi des chiffres non encore publiés conformes aux lignes directrices établies.

### **9.1 Lignes directrices pour l'arrondissement**

Afin que les estimations calculées d'après ces fichiers de microdonnées en vue d'être publiées ou diffusées de toute autre façon correspondent à celles produites par Statistique Canada, il est vivement conseillé à l'utilisateur de les arrondir en se conformant aux lignes directrices qui suivent.

- a) Les estimations qui figurent dans un tableau statistique doivent être arrondies à la centaine près par la méthode d'arrondissement classique. Selon cette méthode, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre retenu ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente d'une unité (1) la valeur du dernier chiffre retenu. Par exemple, si l'on veut arrondir à la centaine près de la façon classique une estimation dont les deux derniers chiffres sont compris entre 00 et 49, il faut les remplacer par 00 et ne pas modifier le chiffre précédent (le chiffre des centaines). Si les deux derniers chiffres sont compris entre 50 et 99, il faut les remplacer par 00 et augmenter d'une unité (1) le chiffre précédent.
- b) Les totaux partiels de marge et les totaux de marge des tableaux statistiques doivent être calculés à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près selon la méthode 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 dénominateurs), puis arrondis à une décimale par la méthode d'arrondissement classique. Si l'on veut arrondir une estimation à un seul chiffre décimal par cette méthode et que le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à retenir ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente d'une unité (1) le dernier chiffre à retenir.
- d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs éléments correspondants non arrondis, puis arrondies à leur tour à la centaine près (ou à la décimale près) selon la méthode d'arrondissement classique.
- e) Si, en raison de contraintes d'ordre technique ou autre, on applique une autre méthode que l'arrondissement classique, si bien que les estimations qui seront publiées ou diffusées de toute autre façon diffèrent des estimations correspondantes publiées par Statistique Canada, il est vivement conseillé à

l'utilisateur d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.

- f) Des estimations non arrondies ne doivent pas être publiées ni diffusées d'aucune façon et en aucune circonstance. Des estimations non arrondies donnent l'impression d'être 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'est pas auto-pondéré. Autrement dit, le poids d'échantillonnage n'est pas le même pour toutes les personnes qui font partie de l'échantillon. Même pour produire des estimations simples, y compris des tableaux statistiques ordinaires, l'utilisateur doit employer le poids d'échantillonnage approprié.

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

L'utilisateur ne doit pas non plus perdre de vue qu'en raison du traitement réservé à la zone du poids, certains logiciels ne permettent pas d'obtenir des estimations qui coïncident exactement avec celles de Statistique Canada.

### **9.2.1 Définitions : estimations de type nominal, estimations quantitatives**

Avant d'exposer la façon de totaliser et d'analyser les données de l'ENSP, il est bon de décrire les deux grandes catégories d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites d'après les fichiers 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 :

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

- ☐ Tous les jours
- ☐ À l'occasion
- ☐ Jamais

### Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures de tendance centrale de quantités basées sur tous les membres de la population observée ou sur certains d'entre eux.

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 correspond à une estimation du nombre de personnes qui fument tous les jours.

Exemple de question quantitative :

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

||| Nombre de cigarettes

### **9.2.2 Totalisation d'estimations de type nominal**

On peut obtenir des estimations du nombre de personnes qui possèdent une caractéristique donnée à partir des fichiers de microdonnées en additionnant les poids finaux de tous les enregistrements contenant la ou les caractéristiques voulues. Pour obtenir les proportions et les rapports de la forme  $\hat{X} / \hat{Y}$ , on doit :

- a) additionner les poids finaux des enregistrements contenant la caractéristique voulue pour le numérateur ( $\hat{X}$ );
- b) additionner les poids finaux des enregistrements contenant la caractéristique voulue pour le dénominateur ( $\hat{Y}$ );
- c) diviser l'estimation du numérateur par celle du dénominateur.

### **9.2.3 Totalisation d'estimations quantitatives**

Pour obtenir des estimations de quantités à partir des fichiers de microdonnées, on doit :

- a) multiplier, pour chaque enregistrement, la valeur de la variable étudiée par le poids final, puis additionner cette quantité pour tous les enregistrements étudiés pour obtenir le numérateur ( $\hat{X}$ );
- b) additionner les poids finaux des enregistrements présentant la caractéristique étudiée pour obtenir le dénominateur ( $\hat{Y}$ );

- c) diviser 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 variable *SMC8\_4*<sup>4</sup> (question SMK-Q4) par le coefficient de pondération WT68, puis on additionne ce résultat pour tous les enregistrements dont la valeur de variable *SMC8\_2* (question SMK-Q2) est «tous les jours» pour obtenir le numérateur ( $\hat{X}$ ). Puis, on additionne les poids finaux de tous les enregistrements dont la valeur de la variable *SMC8\_2* (question SMK-Q2) est «tous les jours» pour obtenir le dénominateur ( $\hat{Y}$ ). Enfin, on divise ( $\hat{X}$ ) 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 (ENSP) se fonde sur un plan de sondage complexe qui prévoit une stratification et un échantillonnage à plusieurs degrés ainsi que la sélection des répondants avec probabilités inégales. L'utilisation des données provenant d'une enquête aussi complexe pose des difficultés aux analystes car le choix des méthodes d'estimation et de calcul de la variance dépend du plan de sondage et des probabilités de sélection.

Nombre de méthodes d'analyse intégrées aux progiciels statistiques permettent d'utiliser des poids, mais la signification et la définition de ces poids diffèrent de celles applicables dans le contexte d'une enquête par sondage. Par conséquent, si les estimations calculées au moyen de ces progiciels sont souvent exactes, les variances n'ont, quant à elles, pratiquement aucune signification.

Dans le cas de nombreuses méthodes d'analyse (par exemple la régression linéaire, la régression logistique, l'analyse de variance), on peut rendre l'application des progiciels courants plus significative en rééchantillonnant les poids qui figurent dans les enregistrements de façon à ce que le poids moyen soit égal à un (1). Les résultats produits par les progiciels classiques sont ainsi plus raisonnables puisque, même s'ils ne reflètent toujours pas la stratification et la mise en grappe du plan de sondage, ils tiennent compte de la sélection avec probabilités inégales. On peut effectuer cette transformation en utilisant dans l'analyse un poids égal au poids original divisé par la moyenne des poids originaux pour les unités échantillonnées (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 coefficients de variation (c.v.) approximatifs pour l'ENSP. On peut utiliser ces tableaux afin d'obtenir des c.v. approximatifs pour les

---

<sup>4</sup> Voir la section 12.2 pour la convention utilisée pour nommer les variables



estimations et les proportions de type nominal. Pour plus de détails, voir le chapitre 10.

#### **9.4 Lignes directrices pour la diffusion**

Avant de diffuser et/ou de publier des estimations tirées de ces 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-après.

##### **Lignes directrices relatives à la variabilité d'échantillonnage**

<b>Type d'estimation</b>	<b>c.v. (en %)</b>	<b>Lignes directrices</b>
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 coefficients de variation approximatifs

Afin de permettre aux utilisateurs d'avoir facilement accès à des coefficients de variation (c.v.) qui s'appliqueraient à une multitude d'estimations de type nominal obtenues à partir de ces fichiers de microdonnées, Statistique Canada a produit un ensemble de Tableaux de coefficients de variation approximatifs (tableaux de c.v.). Ces «tableaux à consulter» permettent aux utilisateurs d'obtenir un c.v. approximatif selon la taille de l'estimation calculée à partir des données de l'enquête.

Les 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 quatre tableaux suivants montrent les effets du plan, la taille des échantillons et les chiffres de population, qui ont servi à produire les Tableaux de coefficients de variation approximatifs.

### **Données d'entrée des tableaux de coefficients de variation, selon les provinces, Canada, membres du ménage, tous les âges**

<b>Province</b>	<b>Effet du plan</b>	<b>Taille de l'échantillon</b>	<b>Population</b>
Terre-Neuve	1,42	2 871	538 051
Île-du-Prince-Édouard	1,46	2 755	135 015
Nouvelle-Écosse	1,42	2 977	910 115
Nouveau-Brunswick	1,45	2 959	734 217
Québec	1,84	8 305	7 157 967
Ontario	1,55	13 526	11 261 692
Manitoba	1,88	3 359	1 069 459
Saskatchewan	1,40	3 092	971 067
Alberta	1,45	4 541	2 846 030
Colombie-Britannique	1,43	4 661	3 900 651
<b>CANADA</b>	1,94	49 046	29 524 264

**Données d'entrée des tableaux de coefficients de variation  
selon le groupe d'âge, Canada, membres du ménage**

Groupe d'âge	Effet du plan	Taille de l'échantillon	Population
0-11	2,55	9 698	4 608 137
12-24	2,36	9 126	5 203 074
25-44	2,14	15 590	9 548 249
45-64	1,78	9 904	6 677 191
65+	1,76	4 728	3 487 613

**Données d'entrée des tableaux de coefficients de variation,  
Selon les provinces et le Canada, personnes sélectionnées**

Province	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve	0,98	963	538 051
Île-du-Prince-Édouard	0,92	932	135 015
Nouvelle-Écosse	0,95	1 071	910 116
Nouveau-Brunswick	1,01	1 073	734 217
Québec	1,14	2 946	7 157 967
Ontario	1,22	4 691	11 261 691
Manitoba	1,19	1 184	1 069 459
Saskatchewan	0,97	1 131	971 067
Alberta	1,03	1 568	2 846 030
Colombie-Britannique	1,01	1 685	3 900 651
<b>CANADA</b>	1,53	17 244	29 524 264

**Données d'entrée des tableaux de coefficients de variation  
selon le groupe d'âge, Canada, personnes sélectionnées**

Groupe d'âge	Effet du plan	Taille de l'échantillon	Population
0-11	1,65	1 995	4 608 137
12-24	1,67	2 526	5 203 074
25-44	1,77	5 775	9 548 249
45-64	1,76	4 097	6 677 191
65+	1,71	2 851	3 487 613

Tous les c.v. présentés dans les tableaux de c.v. sont *approximatifs et ils ne doivent donc pas être considérés comme des valeurs officielles*. Il est possible d'obtenir de Statistique des estimations de la variance réelle de variables particulières. L'utilisation d'estimations de la variance réelle peut permettre aux utilisateurs de diffuser des estimations qui ne pourraient

être diffusées autrement, c'est-à-dire des estimations dont les coefficients de variation se situent dans la catégorie «inacceptable». Consultez la section 10.7 pour plus de détails.

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 c.v. 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 c.v. les coefficients de variation approximatifs pour les 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 coefficients de variation approximatif correspondant à la région géographique appropriée, il faut repérer la valeur de l'estimation 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 c.v. 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 :

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

où  $\hat{X}_1$  représente l'estimation 1,  $\hat{X}_2$  l'estimation 2 et  $\alpha_1$  et  $\alpha_2$  sont les coefficients de variation de  $\hat{X}_1$  et  $\hat{X}_2$  respectivement. Le coefficient de variation de  $\hat{d}$  est donné par  $\sigma_{\hat{d}} / \hat{d}$ . Cette formule donne un résultat exact pour ce qui est de la différence entre des caractéristiques distinctes et non corrélées, mais elle ne donne que des résultats approximatifs dans les autres cas.

### **Règle 4 : Estimations de rapports**

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

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

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

où  $\alpha_1$  et  $\alpha_2$  sont les coefficients de variation de  $\hat{X}_1$  et  $\hat{X}_2$  respectivement.

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

### **Règle 5 : Estimations des différences entre des rapports**

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 5 690 625 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 sélectionnées ;
- 2) La valeur de l'estimation (5 690 625) 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 1,8 % ;
- 4) Le coefficient de variation approximatif de l'estimation est donc 1,8 %. Par conséquent, l'estimation selon laquelle 5 690 625 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  $5\,690\,625 / 6\,622\,143 = 85,9\%$  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 sélectionnées ;
- 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 (85,9 %) et la partie numérateur du pourcentage (5 690 625) pour déterminer le coefficient de variation ;
- 3) Le numérateur (5 690 625) 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 0,7 %, est le coefficient de variation (exprimé en pourcentage) à employer ;
- 5) Le coefficient de variation de l'estimation est donc 0,7 %. Par conséquent, l'estimation selon laquelle 85,9 % 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, 4 738 935/5 690 625 = 83 % 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, 4 661 275 / 6 287 675 = 74 % 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 valeurs des variables SMK8\_2, SMK8\_4, SMK8\_4A, SMK8\_5 et SMK8\_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 pour les personnes sélectionnées, utilisé de la même façon que dans l'exemple 2, vous établissez à 0,7 % le c.v. de l'estimation 1 (exprimé en pourcentage) et à 1,3 % le c.v. de l'estimation 2 (exprimé en pourcentage) ;
- 2) Selon la règle 3, l'erreur-type pour une différence ( $\hat{d} = \hat{X}_2 - \hat{X}_1$ ) est :

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

où  $\hat{X}_1$  est l'estimation 1,  $\hat{X}_2$  est l'estimation 2 et  $\alpha_1$  et  $\alpha_2$  sont les coefficients de variation de  $\hat{X}_1$  et  $\hat{X}_2$  respectivement ;

L'erreur-type de la différence  $\hat{d} = (0,74 - 0,83) = 0,09$  est donc :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[f(0,83)(0,007)]^2 + [f(0,74)(0,013)]^2} \\ &= 0,011\end{aligned}$$

- 3) Le coefficient de variation de  $\hat{d}$  est donné par  $\sigma_{\hat{d}} / \hat{d} = 0,011/0,09 = 0,122$  ;
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 12,2 % (exprimé en pourcentage). Par conséquent, cette estimation peut être publiée sans réserve.

#### **Exemple 4 : Estimations de rapports**

Supposons qu'un utilisateur estime à 5 690 625 le nombre de personnes qui fument tous les jours et à 931 518 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 sélectionnées ;
- 3) Le numérateur de cette estimation de rapport est 931 518. 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 5,0 % ;
- 4) Le dénominateur de cette estimation de rapport est 5 690 625. 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 1,8 % ;
- 5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, soit

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

où  $\alpha_1$  et  $\alpha_2$  sont les coefficients de variation de  $\hat{X}_1$  et  $\hat{X}_2$  respectivement.



C'est-à-dire,

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0,050)^2 + (0,018)^2} \\ &= 0,053\end{aligned}$$

Le rapport des personnes qui fument occasionnellement à celles qui fument tous les jours est 931 518/5 690 625, soit 0,16 :1. Le coefficient de variation de cette estimation est 5,3 % (exprimé en pourcentage). L'estimation peut donc être diffusée sans réserve.

### **10.3 Utilisation des tableaux de c.v. pour calculer les intervalles 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 coefficient de variation approximatifs, 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 (IC) correspondant :

$$IC_X = [ \hat{X} - z \hat{X} \alpha_{\hat{X}}, \hat{X} + z \hat{X} \alpha_{\hat{X}} ]$$

où  $\alpha_{\hat{X}}$  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).

$$\hat{X} = 0,859$$

$$z = 2$$

$\alpha_{\hat{X}} = 0,007$  est le coefficient de variation de cette estimation selon les tableaux.

$$CI_X = \{0,859 - (2) (0,859) (0,007) ; 0,859 + (2) (0,859) (0,007)\}$$

$$CI_X = \{0,847; 0,871\}$$

#### **10.5 Utilisation des 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  $\sigma_{\hat{d}}$ .

Si  $z = (\hat{X}_1 - \hat{X}_2) / \sigma_{\hat{d}}$  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, 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, 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 0,011. Par conséquent,

$$z = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d} = \frac{0,83 - 0,74}{0,011} = \frac{0,09}{0,011} = 8,18$$

Puisque  $z = 8,18$  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 c.v. sont approximatifs, donc non officiels. Statistique Canada peut fournir les coefficients de variation précis pour des estimations particulières (voir la section 12.3 pour plus de détails). Les types d'estimation offerts 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 des fichiers 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, comme 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.-à-d. 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 c.v. à partir des tableaux de coefficients de variation approximatifs 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 le programme de calcul des variances/coefficients de variation réels ou encore soumettre des demandes personnalisées en communiquant avec les Services personnalisés à la clientèle, Division de la statistique de la santé, Statistique Canada au **1-613-951-1746** ou par courriel à l'adresse **hd-ds@statcan.ca**. 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. Les utilisateurs des fichiers de microdonnées à grande diffusion peuvent également faire appel aux services de Téléaccès, sans frais, en envoyant leurs programmes dûment testés par courriel à l'adresse **nphs@statcan.ca** ou en communiquant avec Colette Koeune, Division de la statistique de la santé, Statistique Canada au **1-613-951-1653**.

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

Les quatre 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, pour les membres du ménage et les personnes sélectionnées. Statistique Canada recommande de ne pas publier des estimations basées sur des échantillons de taille inférieure au minimum indiqué dans la colonne «Marginales». Toutefois, si l'utilisateur choisit de le faire, il doit alors indiquer que la qualité des estimations est inacceptable et y adjoindre la lettre I (ou un autre identificateur semblable) avec un avertissement (voir la section 9.4).

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

<b>PROVINCE</b>	<b>ACCEPTABLES</b> (c.v. de 0,0% à 16,5%)	<b>MARGINALES</b> (c.v. de 16,6% à 33,3%)
Terre-Neuve	9 500	2 500
Île-du-Prince-Édouard	2 500	500
Nouvelle-Écosse	15 500	4 000
Nouveau-Brunswick	13 000	3 000
Québec	58 000	14 500
Ontario	47 000	11 500
Manitoba	21 500	5 500
Saskatchewan	16 000	4 000
Alberta	33 000	8 000
Colombie-Britannique	43 500	11 000
<b>CANADA</b>	43 000	10 500

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

<b>GROUPE D'ÂGE</b>	<b>ACCEPTABLES</b> (c.v. de 0,0% à 16,5%)	<b>MARGINALES</b> (c.v. de 16,6% à 33,3%)
0-11	44 000	11 000
12-24	49 000	12 000
25-44	48 000	12 000
45-64	44 000	11 000
65+	47 000	11 500

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

<b>PROVINCE</b>	<b>ACCEPTABLES</b> (c.v. de 0,0% à 16,5%)	<b>MARGINALES</b> (c.v. de 16,6% à 33,3%)
Terre-Neuve	19 500	5 000
Île-du-Prince-Édouard	4 500	1 000
Nouvelle-Écosse	28 500	7 500
Nouveau-Brunswick	24 500	6 000
Québec	100 500	25 000
Ontario	106 500	26 500
Manitoba	38 000	9 500
Saskatchewan	29 500	7 500
Alberta	67 000	17 000
Colombie-Britannique	84 000	21 000
<b>CANADA</b>	96 000	23 500

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

<b>GROUPE D'ÂGE</b>	<b>ACCEPTABLES</b> (c.v. de 0,0% à 16,5%)	<b>MARGINALES</b> (c.v. de 16,6% à 33,3%)
0-11	136 000	34 000
12-24	123 500	31 000
25-44	106 500	26 500
45-64	103 500	26 000
65+	75 000	19 000

## **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 neuf provinces sauf le Québec et un autre pour le Québec. Pour les neuf provinces sauf 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 choisi 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 des échantillons supplémentaires indépendants par composition aléatoire (CA) en Ontario, au Manitoba et en Alberta afin de pouvoir produire des estimations fiables au niveau du secteur sanitaire uniquement pour le cycle 2. En 1998-1999, un échantillon supplémentaire a été tiré pour obtenir un échantillon transversal plus représentatif de la population afin d'améliorer les estimations transversales. Cette meilleure représentativité a été obtenue en compensant d'abord pour l'attrition de l'échantillon puis, ensuite, pour les personnes initialement absentes. Pour compenser l'attrition, tous les logements non-répondants au cycle 1 ont été inclus dans l'échantillon. Pour ce qui est des personnes initialement absentes, ce groupe a été compensé à l'aide d'un sous-échantillon de l'EPA qui incluait des enfants en bas âge ou des nouveaux immigrants, deux sous-populations absentes de l'échantillon original de 1994-1995.

Pour le premier cycle, deux ensembles de facteurs de pondération avaient été 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, trois ensembles de facteurs de pondération ont été nécessaires, tous trois à des fins transversales. Les deux premiers étaient semblables à ceux de 1994-1995 mais étaient calculés différemment pour tenir compte du plan d'échantillonnage de 1996-1997. Le troisième facteur de pondération n'était applicable qu'au fichier de données sur la santé et était requis pour l'analyse de populations et de variables particulières. Au cycle 3, seulement deux ensembles de facteurs de pondération sont nécessaires. Ceux-ci sont semblables à ceux de 1994-1995, mais sont calculés différemment pour tenir compte du plan d'échantillonnage de 1998-1999.

La méthode de pondération de 1998-1999 se fonde en grande partie sur celle de 1994-1995. On prend comme point de départ les facteurs de pondération finaux de 1994-1995 appliqués à chaque répondant longitudinal qui continue à participer à l'enquête. Certaines corrections qui risquent d'être différentes en 1998-1999 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 finaux de 1998-1999, 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 1998-1999. On trouvera une description complète des méthodes appliquées en 1994-1995 dans le Guide des fichiers 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 1998-1999 en ce qui concerne l'échantillon principal de l'ENSP sont décrites à la section 11.1, tandis que les méthodes appliquées en 1998-1999 pour l'échantillon supplémentaire de remise à niveau sont décrites à la section 11.2. Les sections 11.3 et 11.4 décrivent, pour 1994-1995, les procédures relatives aux provinces autres que le Québec et au Québec respectivement.

### **11.1 Pondération de l'échantillon transversal de l'ENSP de 1998-1999, Échantillon principal des ménages**

On décrit à la présente section les méthodes de pondération utilisées pour 1998-1999 et 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 supplémentaires qu'il a fallu utiliser pour tenir compte de l'échantillon supplémentaire de remise à niveau.

#### **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 1998-1999 est le facteur de pondération «épuré», calculé d'après le plan d'échantillonnage original de 1994-1995. Essentiellement, les facteurs de pondération «épurés» sont les simples probabilités inverses de sélection des logements. Une fois ces facteurs de pondération calculés, on effectue les corrections décrites aux sections 11.1.2.1 et 11.1.2.2 pour toutes les provinces à l'exception du Québec pour lequel les corrections sont décrites aux sections 11.1.2.3 et 11.1.2.4.

#### **11.1.2 Corrections aux facteurs de pondération épurés**

##### **11.1.2.1 Corrections pour les provinces autres que le Québec**

###### ***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 sera égal à 1.



Notons que cette correction a dû être recalculée pour tenir compte des quelques logements multiples trouvés dans l'échantillon de remise à niveau pour les logements non-répondants du cycle 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 a tenu compte de l'élimination des unités supplémentaires transversales achetées en 1994-1995, mais qui n'ont pas été étudiées en 1998-1999. On a retranché les ménages sélectionnés en 1994-1995 dans le cadre de l'achat d'unités supplémentaires des calculs. La cinquième correction est effectuée séparément pour l'échantillon longitudinal et celui de remise à niveau. Pour l'échantillon longitudinal, la correction est effectuée comme suit :

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

Cette correction est appliquée au niveau de la strate pour chaque «saison», c'est-à-dire l'été et l'hiver. Les strates de l'ENSP sont ici 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 variait significativement d'une saison à l'autre. Les facteurs de pondération auxquels on fait référence ici correspondent aux facteurs de pondération de base de l'EPA multipliés par les facteurs de correction établis jusqu'à présent soit les corrections 1 à 4. La correction repose sur l'hypothèse que les ménages interrogés présentent les mêmes caractéristiques que ceux qui ne l'ont pas été, et ce, pour chaque groupe strate-saison.

Pour l'échantillon de remise à niveau, la correction est faite séparément pour les ménages non-répondants au cycle 1 et pour le reste de l'échantillon de remise à niveau, c'est-à-dire pour les enfants en bas âge et les nouveaux immigrants. La correction pour les ménages non-répondants au cycle 3 est traitée ici de façon similaire à ce qui a été fait pour l'échantillon longitudinal. Toutefois, à cause des petites tailles d'échantillon, les corrections ont été faites au niveau des provinces/RMR/urbain. Notons que tous les logements

hors champs d'enquête (en construction, ménages rejetés, etc.) sont supprimés.

#### ***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 (c.-à-d. une personne de moins de 25 ans). Ces ménages «rejetés» viennent de la partie de l'échantillon «enfants» épurée d'après la composition du ménage.

On a utilisé cette méthode pour compenser la sur-représentation des ménages de petite taille dans l'échantillon et la sous-représentation des ménages de grande taille, tout en maintenant la sur-représentation voulue pour les individus plus âgés. 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 enfant. Puisqu'on en écarte ou «rejette» certains, les ménages sans jeune ni enfant 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 pour les ménages des trimestres 3 et 4, sans jeune ni enfant 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 aux strates des ménages vivant en appartement, à celles des ménages à revenu élevé, ni à celles des ménages habitant dans une région éloignée, puisque la méthode d'exclusion n'avait pas été employée pour ces trois strates.

Notons que les logements non-répondants du cycle 1 provenant de l'échantillon de remise à niveau ont été inclus pour cette correction.

#### **11.1.2.2 Corrections supplémentaires pour les membres du panel, pour les provinces autres que le Québec**

Les facteurs de pondération épurés ajustés sont obtenus en multipliant les facteurs de pondération épurés par les facteurs de correction 4, 5, 6, puis ensuite par les facteurs de correction 7B, 8B et 9B décrits ci-dessous.

Un membre de chaque ménage répondant a été sélectionné pour devenir membre du panel de l'étude longitudinale. Si la personne sélectionnée est 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»), alors 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) en 1994-1995. Dans les autres cas, on a posé au membre du panel une 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 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, tous les enfants de moins de 12 ans, jusqu'à concurrence de quatre, étaient sélectionnés et répondaient 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 de l'ENSP. Les données sur les enfants ne figuraient pas dans les fichiers 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 et pour les cycles subséquents, 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 y ont contribué en 1998-1999, 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.

Notons que les logements non-répondants du cycle 1 provenant de l'échantillon de remise à niveau ont été inclus pour cette correction.

***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.

Notons que la même correction a été faite pour l'échantillon de remise à niveau des logements non-répondants du cycle 1.

***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.

Dans les ménages avec jeunes mais sans enfant, 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 enfant 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 jeune ni enfant, 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.

En raison de cette même erreur, notons qu'il a été nécessaire d'inclure pour cette correction les logements non-répondants du cycle 1. Pour la partie de l'échantillon de remise à niveau provenant de l'EPA, les membres sélectionnés avaient été présélectionnés à partir de leurs dernières interviews avec l'EPA. Leur sélection n'a donc pas été affectée par cette erreur.

Les «facteurs de pondération» susmentionnés sont les facteurs de pondération épurés multipliés par tous les facteurs de correction calculés jusqu'à présent (c'est-à-dire les corrections 4 à 6, ainsi que 7B, 8B et 9B). Ils représentent les facteurs de pondération épurés «ajustés» pour les provinces autres que le Québec et sont utilisés pour créer les facteurs de pondération relatifs au ménage et au membre sélectionné.

### **11.1.2.3 Corrections pour le Québec**

#### ***Correction 4Q : 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.

Notons que cette correction a dû être recalculée pour tenir compte des quelques logements multiples trouvés dans l'échantillon de remise à niveau pour les logements non-répondants du cycle 1.

***Correction 4BQ: Correction pour l'expansion de la grappe***

Dans quelques cas, un relistage des grappes a été nécessaire pour l'ENSP. 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 :

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

Cette correction a été appliquée à chaque logement sélectionné dans la grappe. On a supposé que les écarts inférieurs à 15% étaient négligeables et, dans ces 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 :

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

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 5Q: Correction pour la non-réponse du ménage***

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

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

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 de réponse de l'ESS était 10 (démoli, vide, abandonné). Le second comprend les logements dont le code de 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.

Pour la partie de l'échantillon de remise à niveau composée des ménages non-répondants au cycle 1, tous les ménages non-répondants au cycle 3 sont traités de façon similaire. Notons que tous les logements hors du champs de l'enquête (en construction, ménages rejetés, etc.) sont supprimés.

#### **11.1.2.4 Corrections supplémentaires pour les membres du panel pour le Québec**

Un membre de chaque ménage répondant a été sélectionné pour devenir membre du panel de l'étude longitudinale. Si la personne sélectionnée est 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») alors 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) en 1994-1995. Dans

les autres cas, on a posé au membre du panel une 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 7BQ : 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 les fichiers 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 enfant 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 et elle est expliquée plus loin à la partie concernant la correction pour les enfants de 12 ans.

Notons que les logements non-répondants du cycle 1 provenant de l'échantillon de remise à niveau ont été exclus pour cette correction.

***Correction 8BQ : Probabilité inverse de sélection des membres du panel***

En principe, dans les logements de l'échantillon «enfants» où on n'a pas recensé d'enfant 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.

Notons que la même correction a été faite pour l'échantillon de remise à niveau des logements non-répondants du cycle 1. Pour celui provenant de l'EPA, chaque membre du ménage «initialement absent» (c'est-à-dire les enfants en bas âge et immigrants) recevait une probabilité de sélection égale, basée sur la liste des membres du ménage lors de la dernière interview pour l'EPA.

### ***Correction 9BQ : 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 de 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.

Notons aussi qu'à cause de cette même erreur, il a été nécessaire d'inclure pour cette correction, les logements non-répondants du cycle 1 de l'échantillon de remise à niveau. Puisque les personnes sélectionnées pour l'échantillon de remise à niveau provenant de l'EPA avaient été présélectionnées à partir de la dernière interview de leur ménage avec l'EPA, leur sélection n'a pas été affectée par cette erreur.

Les facteurs de pondération auxquels on réfère ici sont les facteurs de pondération épurés multipliés par tous les facteurs de correction mentionnés jusqu'à maintenant (c.-à-d. les corrections 4Q, 4BQ, 5Q, 7BQ, 8BQ et 9BQ). Ceux-ci représentent les facteurs de pondération épurés ajustés pour le Québec et sont utilisés pour créer les facteurs de pondération relatifs au ménage et au membre sélectionné.

### **11.1.3 Corrections des facteurs de pondération appliqués aux membres des ménages**

#### ***Correction 10A : 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 1998-1999 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 les fichiers 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 :

*somme des poids pour les ménages répondants et non – rép. dans la catégorie de pondération*

*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 troisième 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 ANGROSS 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)<sup>5</sup>.

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 1998-1999 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 :

---

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

<b>Géographiques</b>	Province, région métropolitaine de recensement, indicateur de région urbaine/rurale
<b>Ménage</b>	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.
<b>Personnel</b>	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.

Notons que la non-réponse au niveau des ménages pour l'échantillon de remise à niveau des logements non-répondants du cycle 1 a été traitée avec la correction 5. L'échantillon de remise à niveau provenant de l'EPA a été ajusté séparément utilisant seulement des classes de base à cause des petites tailles d'échantillon.

#### ***Correction 11A : 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 ne sera associé aux autres membres du ménage observés au troisième cycle, puisque ceux-ci n'avaient pas été sélectionnés dans le panel au premier cycle. La méthode du partage des facteurs de pondération permet d'attribuer des facteurs de pondération à ces personnes de façon à produire des estimations non biaisées.

Au deuxième cycle, on a attribué à chaque membre du ménage un facteur de pondération égal à celui 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 qui sont nées ou sont entrées au Canada depuis 1994). Cependant, pour le troisième cycle, comme on a tenu compte des personnes absentes au départ grâce au tirage de l'échantillon de remise à niveau d'immigrants et de nouveau-nés de l'EPA, il est inutile de «donner» un coefficient de pondération supplémentaire à ces nouveaux membres du panel. Par conséquent, on applique une version simplifiée de la méthode du partage des facteurs de pondération consistant à partager le facteur de pondération également entre tous les membres du ménage. Autrement dit, on attribue à chaque membre du ménage le facteur de pondération du membre du panel divisé par le nombre *total* de membres du ménage qu'ils aient fait partie ou non du ménage au premier cycle.

Tel que mentionné dans la section 5.1.2, dû à une erreur dans l'application informatique de collecte des données, aucun enfant de 0 à 11 ans n'a été sélectionné comme membre du panel longitudinal dans l'échantillon de remise à niveau de logements non-répondants au premier cycle. Pour cette population, le facteur de pondération du membre du panel doit être *donné* à tous les membres du ménage de 0 à 11 ans et *partagé* avec les autres membres. Autrement dit, dans ces ménages, on attribue à chaque membre le facteur de pondération du membre du panel divisé par le nombre de membres du ménage ayant plus de 11 ans. Pour plus de renseignements, consulter Ernst (1989)<sup>6</sup> ou Lavallée (1995)<sup>7</sup>.

#### ***Correction 12A : Migrations inter provinciales***

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 inter provinciale observé. Notons que les corrections ont aussi été faites pour certains individus n'ayant pas déménagé pour réduire les poids extrêmes.

#### ***Correction 13A : 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 l'on fait la correction qui suit, par province :

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

On notera que la correction pour les ménages sélectionnés dans l'échantillon de remise à niveau a été faite séparément en se servant de catégories très élémentaires.

#### ***Correction 14A : Intégration de l'échantillon principal et des échantillons de remise à niveau***

Comme on l'a mentionné précédemment, au troisième cycle on a sélectionné un échantillon supplémentaire de remise à niveau. Cependant, contrairement au

---

<sup>6</sup> 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.

<sup>7</sup> Lavallée, P. (1995). Cross-sectional Weighting of Longitudinal Surveys of Individuals and Households Using the Weight Share Method. *Survey Methodology*, 21, 25-32

deuxième cycle, où certaines provinces ont acheté des unités d'échantillonnage supplémentaires, il n'est pas nécessaire de procéder à l'intégration explicite de cet échantillon de remise à niveau dans l'échantillon principal. L'échantillon de remise à niveau de logements non-répondants au premier cycle est entièrement intégré à l'échantillon principal puisqu'il existe des facteurs de pondération de base «épurés» du premier cycle pour les unités qui le constituent. L'autre partie de l'échantillon de remise à niveau, c'est-à-dire l'échantillon de ménages comptant des nouveau-nés ou des nouveaux immigrants, peut être considérée comme un échantillon de sous-populations ou de strates conceptuellement différentes de la population de base. Cette dernière et les «strates» de remise à niveau sont mutuellement exclusives et exhaustives en ce qui concerne la population du Canada et peuvent donc être simplement additionnées aux fins de l'estimation. Il ne reste donc plus qu'à étalonner les poids de sondage en prenant pour référence les totaux nationaux de population.

#### ***Correction 15A : Correction par stratification a posteriori***

Cette correction fait en sorte que la somme des facteurs de pondération finaux est égale aux totaux de population de 1998-1999 au niveau province-âge-sexe. Ces totaux sont établis d'après les projections post censitaires de 1996. Les catégories âge-sexe comptent 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. Cette correction est donnée par

*projection démographique pour une catégorie province - âge - sexe*

---

*somme des facteurs de pondération des répondants dans une catégorie province - âge - sexe*

On notera que les ménages entièrement composés d'immigrants arrivés en 1994 ou après sont maintenant couverts par l'échantillon transversal grâce à l'échantillon de remise à niveau de ménages d'immigrants et sont donc inclus dans les projections démographiques.

#### ***Correction 16A : Facteur de bruit***

Pour des raisons de confidentialité, on ajoute un facteur de «bruit» aux facteurs de pondération des personnes qui font partie d'un même ménage. On choisit ce facteur, qui obéit à une loi de distribution uniforme, de façon à ce 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. Le facteur de pondération résultant est appelé **«facteur de pondération final type pour les membres des ménages» (WT58)**.

#### **11.1.4 Corrections de la pondération pour les membres du panel**

Pour calculer le facteur de pondération final appliqué aux membres du panel dans le Fichier 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 10B : 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 10A, on définit 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 cycles précédents, comme cela est le cas pour le calcul de la correction pour la non-réponse des membres du ménage. La correction est donnée par :

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

À noter que la correction 5 tient compte de la non-réponse des ménages sélectionnés dans l'échantillon de remise à niveau de ménages non-répondants au premier cycle. En raison de la petite taille des échantillons, on a procédé à des corrections distinctes pour tenir compte de la non-réponse des membres du panel sélectionnés dans ces ménages ainsi que dans l'échantillon de remise à niveau de ménages de l'EPA.

##### ***Correction 12B : Migrations inter provinciales***

Cette correction est similaire à la correction 12A. Quand on observe un facteur de pondération extrême pour le schéma de migration inter provinciale observé, on réduit les facteurs de pondération appliqués à tous les membres concernés du panel de sorte que leur somme soit égale à la projection démographique du nombre de personnes qui ont déménagé au cours des deux dernières années conformément au schéma de migration. On n'applique pas la correction si celle-ci se traduit par une augmentation du facteur de pondération puisqu'elle augmenterait encore davantage la valeur du facteur de pondération extrême.

***Correction 14B : Intégration de l'échantillon principal et des échantillons de remise à niveau***

Comme pour la correction 14A, cette correction n'est pas explicitement nécessaire pour 1998-1999.

***Correction 15B1 : Stratification a posteriori du groupe des 0 à 11 ans***

Alors que la population de 0 à 3 ans est représentée correctement dans l'échantillon de remise à niveau de ménages comptant des nouveau-nés de l'EPA, la population de 4 à 11 ans n'est pas représentée correctement dans le panel longitudinal à cause d'une erreur dans l'application d'application informatique de collecte des données (voir correction 11A). Cette population était représentée dans l'échantillon principal, mais ne l'a pas été dans l'échantillon de remise à niveau. Pour compenser cette sous-représentation, on a procédé à une «pré stratification a posteriori » avant la correction 15B2. On a étalonné séparément les deux groupes d'âge en question, c'est-à-dire celui des 0 à 3 ans et celui des 4 à 11 ans, en prenant pour référence les totaux de population de 1998-1999 au niveau Canada-âge-sexe. Cette correction est donnée par :

$$\frac{\text{projection démographique pour une catégorie âge - sexe}}{\text{somme des facteurs de pondération des répondants dans une catégorie âge - sexe}}$$

***Correction 15B2 : Stratification a posteriori***

Cette correction assure que la somme des facteurs de pondération finaux qui figurent dans le Fichier santé soit égale aux totaux de population de 1998-1999 au niveau province-âge-sexe. Les catégories âge-sexe comptent 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. Cette correction est donnée par :

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

À noter que, grâce à l'échantillon de remise à niveau de ménages comptant des nouveau-nés, le premier groupe d'âge couvre de nouveau entièrement le groupe des 0 à 11 ans, alors qu'en 1996-1997 les enfants de moins d'un an et d'un an ne faisaient pas partie du champ d'observation. De nouveau, il convient de souligner que les ménages composés entièrement d'immigrants arrivés depuis 1994-1995 sont couverts maintenant par l'échantillon de remise à niveau de l'ENSP de 1998-1999 et, donc, inclus dans les projections démographiques. Le facteur de pondération résultant est appelé «**facteur de pondération type final pour les membres du panel** » (WT68).



## **11.2 Pondération transversale de l'échantillon de l'ENSP de 1998-1999, Échantillon supplémentaire de remise à niveau**

Un échantillon supplémentaire a été sélectionné pour améliorer la couverture de la population de 1998-1999 et pour compenser l'érosion de l'échantillon original. Cet échantillon, qui comprend deux fractions, ne servira qu'à la production d'estimations transversales et ne sera utilisé qu'au troisième cycle. (Voir la section 5, Plan d'échantillonnage pour plus de précisions.) Pour tirer pleinement parti des renseignements transversaux que fournissent les données, il est nécessaire de regrouper celles tirées de l'échantillon principal de l'ENSP et celles provenant de l'échantillon supplémentaire de remise à niveau en un grand ensemble et de calculer des facteurs de pondération qui reflètent ce regroupement des sources de données. (Voir la section 12.1, Utilisation des facteurs de pondération pour plus de précisions sur les conditions dans lesquelles doivent être utilisés les divers facteurs de pondération.)

La première fraction de l'échantillon de remise à niveau couvre la part de la population du Canada qui «ne faisait pas partie du champ d'observation de l'enquête au départ» en 1994-1995, soit plus précisément les immigrants arrivés au Canada depuis le 1<sup>er</sup> janvier 1995 et les enfants qui y sont nés depuis cette date. Le sous-échantillon d'immigrants, qui compte 400 ménages d'immigrants, a été sélectionné à l'aide d'une question filtre posée à quatre groupes de rotation de l'Enquête sur la population active. Le sous-échantillon de ménages comptant des enfants nés après le 1<sup>er</sup> janvier 1995, qui comprend environ 760 ménages, a été sélectionné parmi les ménages avec enfants de 0 à 3 ans faisant partie de deux de ces quatre groupes de rotation de l'EPA.

La deuxième fraction de l'échantillon de remise à niveau a été sélectionnée parmi les logements non-répondants qui auraient dû faire partie de l'échantillon original au premier cycle, en vue de compenser l'érosion du panel survenue au cours des quatre années écoulées depuis sa création. Pour cette fraction d'échantillon, plus de 2 800 logements ont été sélectionnés. Cela comprend les logements pour lesquels on n'avait pas obtenu de réponse ou dont les occupants avaient refusé de répondre, ainsi que des logements qui avaient posé des problèmes techniques. On a cependant exclu les logements hors du champ d'observation de l'enquête, comme les immeubles inoccupés.

Les corrections particulières de la pondération de l'échantillon de remise à niveau de l'EPA sont exposées dans la présente section. Les corrections de la pondération applicable aux logements non-répondants au premier cycle sont intégrées à la section 11.1, puisque les facteurs de pondération initiaux ont été calculés au premier cycle en même temps que ceux de tous les autres logements faisant partie de l'échantillon principal de ménages.

### **11.2.1 Facteurs de pondération de base de l'EPA**

Comme pour 1994-1995 (voir la section 11.3.1), le point de départ est le facteur de pondération de base de l'EPA, c'est-à-dire le produit du facteur de pondération de la grappe et du facteur de pondération du logement.

### **11.2.2 Sous-facteurs de pondération de l'EPA**

Comme il fallait connaître la composition du ménage lors de la dernière entrevue de l'EPA pour sélectionner l'échantillon de remise à niveau de l'EPA, seuls les ménages répondants ont pu être pris en considération. Pour compenser cette réduction de la taille réelle de l'échantillon, on a calculé les sous-facteurs de pondération de l'EPA que l'on a ensuite légèrement gonflés pour tenir compte de la non-réponse.

### **11.2.3 Corrections supplémentaires des facteurs de pondération en fonction des sous-facteurs de pondération**

#### ***Correction 1T : Correction du facteur de pondération du groupe de rotation***

L'échantillon complet de l'EPA comprend six «groupes de rotation». Puisque l'échantillon de remise à niveau de ménages comptant des nouveau-nés ne se fonde que sur deux groupes de rotation, on applique au facteur de pondération un facteur multiplicatif de correction de 6/2. De la même façon, on applique un facteur de correction de 6/4 dans le cas de l'échantillon de remise à niveau de ménages d'immigrants.

#### ***Correction 2T : Correction du facteur de pondération du sous-échantillon de ménages comptant des nouveau-nés***

Cette partie de l'échantillon de remise à niveau se fonde sur deux groupes de rotation de l'EPA qui ont produit un plus grand nombre que nécessaire de ménages comptant des nouveau-nés. Par conséquent, on sélectionne environ la moitié de ces ménages. On applique un poids d'échantillonnage plus élevé aux ménages qui comptent plus d'un nouveau-né qu'à ceux qui n'en comptent qu'un seul afin de stabiliser les facteurs de pondération au niveau du membre du panel.

#### ***Correction 3T : Correction du facteur de pondération pour les nouveau-nés de l'ELNEJ***

Dans le deuxième des deux groupes de rotation de l'EPA susmentionné, les ménages comptant des nouveau-nés (enfants de moins de 12 mois) ont été répartis entre l'ENSP et l'ELNEJ. On a procédé à une deuxième correction presque égale à 2 pour les ménages visés afin de tenir compte de ce sous-échantillonnage. Cette correction a été appliquée uniquement pour le deuxième groupe de rotation utilisé durant le troisième trimestre de collecte des données de l'ENSP.

### ***Correction 8T : Probabilité de sélection inverse du membre du panel***

Dans le cas de l'échantillon de remise à niveau de l'EPA, on a attribué à chaque membre du ménage «ne faisant pas partie du panel au départ» (c.-à-d. les nouveau-nés et les nouveaux immigrants) une probabilité égale de sélection calculée d'après la liste de membres du ménage lors de la dernière entrevue de l'EPA.

Pour l'échantillon de remise à niveau de l'EPA, le «facteur de pondération du membre du panel» est le facteur de pondération de base de l'EPA multiplié par toutes les corrections effectuées jusqu'à ce stade (c.-à-d. la correction pour la non-réponse à l'EPA ainsi que les corrections 1T, 2T, 3T et 4T). On se sert de ce facteur de pondération comme donnée pour le calcul des facteurs de pondération des membres du ménage et du membre du panel. (Voir les sections 11.1.3 et 11.1.4.)

## **11.3 Méthodes de pondération fondées sur celles de 1994-1995 pour les provinces sauf le Québec**

On présente d'abord une brève description de la méthode de pondération fondamentale de l'EPA qui inclut les facteurs de pondération multiplicatifs initiaux utilisés pour calculer les facteurs de pondération «épurés». Notons que pour la création des facteurs de pondération épurés, on a dû remonter plus loin dans le processus pour le cycle 3 que ce qui avait été fait pour le cycle 2, c'est-à-dire que l'on a remonté jusqu'à la correction 3. Ceci était nécessaire à cause de l'échantillon de remise à niveau transversal qui incluait les logements non-répondants du cycle 1. La correction 4 et toutes celles qui suivent ont dû être revues pour réintégrer correctement ces logements dans la structure de pondération.

### **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). Dans chaque cas 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 rotation***

L'échantillon de l'EPA se compose de six «groupes de rotation». Pour l'ENSP, on prélève un échantillon de l'EPA en choisissant un nombre entier (de 1 à 6) de groupes de rotation. 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 :

$$\frac{\text{nombre de groupes de rotation dans une strate de l'EPA utilisés pour l'EPA (habituellement 6)}}{\text{nombre total de groupes de rotation dans une strate de l'EPA nécessaires pour l'ENSP}}$$

#### ***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 a instauré un sous-échantillonnage analogue des grappes qui affichaient 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 rotation de l'EPA dont on n'a pas besoin pour l'ENSP (voir correction pour le groupe de rotation). La troisième correction s'effectue comme suit :

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

Les facteurs de pondération référés ci-dessus sont les facteurs de pondération de base de l'EPA multipliés par tous les facteurs de correction mentionnés jusqu'ici c'est-à-dire, les facteurs de correction 1 à 3. Ceux-ci représentent les facteurs de pondération «épurés» pour les provinces autres que le Québec et sont le point de départ pour la pondération de 1998-1999.

#### **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 1998-1999 des membres de l'ENSP. Notons encore une fois que pour obtenir les facteurs de pondération épurés on a dû, dans le processus, remonter plus loin au cycle 3 que ce qui avait été fait pour le cycle 2, c'est-à-dire qu'on a dû remonter jusqu'au facteur de pondération de base du logement. Ceci était nécessaire à cause de l'échantillon de remise à niveau transversal qui incluait les logements non-répondants du cycle 1. Toutes les corrections mentionnées à partir de ce point ont dû être revues pour réintégrer de façon adéquate ces logements dans la structure de pondération.

##### **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 soit un relèvement de 100% ou de 40% de la probabilité d'être retenues.

##### ***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.

Le facteur de pondération «épuré» pour 1998-1999 est le facteur de pondération de base du logement et représente le point de départ pour la pondération de 1998-1999.

## **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 à partir des 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, Fichier général transversal, WT58**

Un seul facteur de pondération, WT58, 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 l'échantillon total, c'est-à-dire sur l'intégration des échantillons principal et supplémentaire. 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, Fichier sur la santé transversal, WT68**

Un seul facteur de pondération, WT68, figure dans le fichier sur la santé. Ce facteur de pondération s'applique à tous les groupes et à toutes les provinces. Il est fondé sur l'échantillon total, c'est-à-dire sur l'intégration des échantillons principal et supplémentaire. ON DEVRAIT SE SERVIR DE CE FACTEUR DE PONDÉRATION POUR ANALYSER TOUTES LES QUESTIONS AYANT TRAIT AU FICHIER SUR LA SANTÉ.

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.

### **12.2 Convention utilisée 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 : d'une part limiter les noms des variables à huit caractères au plus pour qu'il soit facile de les utiliser avec les logiciels d'analyse et d'autre part, 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 recueillies en 1994-1995, en 1996-1997 et en 1998-1999. Les noms des variables pour ces questions 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 recueillies. Cette convention sera appliquée pendant toute la durée de l'enquête longitudinale et sera adoptée pour toutes les composantes de l'ENSP, soit la composante des ménages, la composante des établissements de santé, 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 et DHC8\_AGE est la suivante :

<b>DH :</b>	dans la section du questionnaire sur la démographie et la composition du ménage ;
<b>C :</b>	questions qui font partie de la composante principale de l'enquête auprès des ménages ;
<b>4 :</b>	figure dans le cycle de 1994-1995 ;
<b>6 :</b>	figure dans le cycle de 1996-1997 ;
<b>8 :</b>	figure dans le cycle de 1998-1999 ;
<b>_ :</b>	peut être trouvé sur le questionnaire ;
<b>AGE :</b>	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	IN	Revenu
AL	Alcool	IS	Assurance
AM	Administration (de l'enquête)	LF	Population active
AP	Opinions au sujet des parents	MH	Santé mentale
BP	Tension artérielle	NU	Nutrition
CC	Problèmes de santé chroniques	PA	Activités physiques
CE	Abandon du contact	PC	Examen général
CI	Renseignement sur le contact (établissements, 1998-1999)	PR	Province
DG	Consommation de médicaments	PY	Ressources psychologiques (estime de soi, contrôle, sentiment de cohésion) (Cycles 1 et 3)
DH	Démographie et ménage	RA	Limitation des activités
DV	Visites chez le dentiste	RP	Mouvements répétitifs
ED	Niveau de scolarité	RS	Sécurité routière
ES	Services d'urgence	SC	Soins personnels
EX	Examen de la vue	SD	Renseignements socio-démographiques
FH	Antécédents médicaux de la famille	SH	Santé sexuelle
FI	Précarité alimentaire	SM	Usage du tabac
FL	Équilibre et chutes (établissements)	SP	Identificateurs d'échantillon (méthodologie)
FS	Vaccination antigrippale	SS	Soutien social
GE	Identificateurs géographiques (méthodologie)	ST	Stress
GH	État de santé général	TA	Variantes de l'usage du tabac
HC	Utilisation des soins de santé	TU	Bronzage et exposition aux rayons UV
HH	Ménage	TW	Incapacité au cours des deux dernières semaines
HI	Renseignements sur la santé	WH	Santé de la femme : auto-examen des seins, examen des seins, mammographie et test de Papanicolaou
HS	État de santé	WT	Facteurs de pondération
HV	VIH		
HW	Taille et poids		
IJ	Blessures		

### **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 de santé
- 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 Suppléments nationaux (Enquête promotion santé et Supplément sur la précarité alimentaire parrainé par DRHC en 1998)
- 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
- 5 1996-1997
- 6 1998-1999
- 0 2000-2001
- 2 2002-2003
- A 2004-2005
- B 2006-2007
- C 2008-2009
- D 2010-2011
- E 2012-2013
- F 2014-2015

### 12.2.5 Position 5 : Type de variable

–	Variable recueillie	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 recueillies (comme une variable dérivée), mais ordinairement par l'application informatique de collecte des données, 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 : ainsi 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 : ainsi 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 DHC8\_AGE en 1998-1999. Dans les cas où les réponses à des groupes de questions portant sur le même sujet ont été recueillies 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 Accès aux données des fichiers maîtres**

Afin de respecter le droit à la vie privée des personnes qui participent à l'enquête, il faut s'assurer que les fichiers de microdonnées répondent aux normes strictes de sécurité et de confidentialité prévues par la *Loi sur la statistique* avant de permettre aux membres du public de les consulter. 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 maîtres de l'enquête est un moyen de récolter ces fruits. On peut fournir à l'acheteur d'un produit de microdonnées un fichier maître «fictif» 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 fictif pour confirmer que les routines fonctionnent convenablement. Il ne reste plus qu'à envoyer le programme pour les totalisations personnalisées par courriel à [nphs@statcan.ca](mailto:nphs@statcan.ca). Le programme de totalisation est transmis au réseau interne protégé de Statistique Canada et appliqué au fichier maître de données de l'ENSP approprié. Les données de sortie sont filtrées pour s'assurer qu'elles sont conformes aux normes de confidentialité et de fiabilité puis sont transmises 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 de la statistique de la santé de produire des totalisations personnalisées (courriel [hd-ds@statcan.ca](mailto:hd-ds@statcan.ca)). Ce service permet aux utilisateurs d'obtenir des totalisations à partir du fichier maître sans soumettre de programme informatique. 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.

Finalement, un programme de recherche permet aux chercheurs de soumettre à Statistique Canada, un projet de recherche utilisant les données des fichiers maîtres de l'ENSP. L'acceptation de ces projets de recherche suit un ensemble de règles bien précises. Lorsque le projet est accepté, le chercheur devient employé réputé de STC et peut accéder les données des fichiers maîtres à partir des sites désignés de Statistique Canada. Pour plus d'information au sujet de ce programme veuillez contacter Mario Bédard par téléphone au 1-613-951-8933, par télécopie au 1-613-951-4198 ou par courriel à l'adresse suivante : [nphs@statcan.ca](mailto:nphs@statcan.ca).